

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:06:59 ; Search time 24.81 seconds  
(without alignments)  
1385.329 Million cell updates/sec

Title: US-09-426-072-2  
Perfect score: 2501  
Sequence: 1 MKSFALTGLGLASIGLAS.....GTTLRDAPITKTPHTFSV 464

Scoring table: BLOSUM62  
Gapop 10.0, Capext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2501	100.0	464	22	Phospholipase B of
2	126.5	5.1	483	11	Secretory signal P
3	111.5	4.5	1041	18	ATH1 gene product,
4	109.5	4.4	579	14	PSI875 swine parvo
5	109.5	4.4	925	21	Agarase 1-7 protei
6	109.5	4.4	1041	20	S. cerevisiae acid
7	108.5	4.3	626	21	Arabidopsis thalia
8	107	4.3	532	21	Adeno associated v
9	107	4.3	588	21	Adeno associated v
10	107	4.3	724	21	Adeno associated v
11	107	4.3	1257	13	Bacillus thuringie

12	107	4.3	1257	13	AAR28811	BT toxin 33F2. Ba
13	107	4.3	1257	13	AAR29518	BT toxin 33F2. Ba
14	107	4.3	1257	14	AAR44203	Bacillus thuringie
15	107	4.3	1257	18	AAW13886	33F2 toxin. Bacil
16	107	4.3	1257	21	AAB13893	Bacillus thuringie
17	107	4.3	1257	22	AAB59883	33F2 protein. Bac
18	106	4.2	1257	13	AAR20068	B thuringiensis to
19	105.5	4.2	962	21	AAB18211	Plasmodium falci
20	105.5	4.2	1041	20	AAI29830	Saccharomyces cere
21	105.5	4.2	1122	16	AAR64927	Cytadhesin protein
22	104.5	4.2	737	16	AAR70186	Arg-gingipain-1.
23	104.5	4.2	737	18	AAW34846	Arg-gingipain high
24	104.5	4.2	737	21	AAW67395	Arg-gingipain-2 am
25	104.5	4.2	991	16	AAR77313	porphyromonas ging
26	104.5	4.2	1687	17	AAW86033	P. gingivalis haem
27	104.5	4.2	1687	19	AAW69495	Haemagglutinin pro
28	104.5	4.2	1704	16	AAW70188	Arg-gingipain-2 pr
29	104.5	4.2	1704	18	AAW34843	Arg-gingipain high
30	104.5	4.2	1704	21	AAW67396	Arg-gingipain-2 am
31	104	4.2	783	21	AAW75533	Neisseria meningit
32	104	4.2	783	21	AAW75534	Neisseria meningit
33	103.5	4.1	414	21	AAW33292	Eucalyptus grandis
34	103.5	4.1	579	14	AAW38702	Swine parvovirus B
35	103.5	4.1	579	17	AAW89721	Swine parvovirus B
36	103.5	4.1	598	8	AAW70500	Pig parvo virus B
37	103.5	4.1	1040	21	AAW30935	Arabidopsis thalia
38	103.5	4.1	1052	21	AAW30934	Arabidopsis thalia
39	103.5	4.1	1181	21	AAW30933	Arabidopsis thalia
40	103.5	4.1	1290	18	AAW07609	Rat von Ebner's gl
41	103	4.1	636	20	AAW89548	Bacillus sp. alk
42	103	4.1	686	5	AAW40068	Sequence of a porc
43	102	4.1	433	13	AAW26274	Alkali-protease ya
44	102	4.1	433	19	AAW61495	Modified Bacillus
45	102	4.1	433	20	AAW95698	Bacillus sp. Lion

ALIGNMENTS

RESULT 1  
AAB04146  
ID AAB04146 standard; Protein; 464 AA.  
XX AAB04146;  
XX 11-APR-2001 (first entry)  
XX Phospholipase B of Aspergillus oryzae.  
DE Phospholipase B: hydrolysis; hydrolyse; lysophospholipid;  
KW Phospholipid; lecithin; lysolecithin; degumming; starch;  
KW wheat starch; oil; dough; baking; strength; elasticity;  
KW stability; extensibility.  
XX Aspergillus oryzae.  
OS  
XX  
XX US6146869-A.  
XX  
XX 14-NOV-2000.  
XX  
XX 21-OCT-1999; 99US-0426072.  
XX  
XX 21-OCT-1999; 99US-0426072.  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
XX Harris P, Brown KM;  
XX WPI; 2001-049057/06.  
XX DR N-PSDB; AAA54535.  
XX Novel polypeptide having phospholipase B activity useful for  
XX hydrolysing fatty acyl groups of phospholipid, preparing dough or baked

PT product and reducing phospholipid content in an edible oil  
 XX Claim 1; Fig 1a-b; 24pp; English.  
 XX  
 CC Phospholipase B is useful for hydrolysing the fatty acyl group(s) of  
 CC phospholipids or lysophospholipids, such as lecithin or lysolecithin,  
 CC degumming an aqueous carbohydrate solution or slurry to improve its  
 CC filterability, particularly, a starch hydrolysate or wheat starch  
 CC hydrolysate which is difficult to filter and to reduce the  
 CC phospholipid content in an edible oil by treating the oil with  
 CC phospholipase B to hydrolyse a major portion of the phospholipid  
 CC and separating an aqueous phase containing the hydrolysed  
 CC phospholipid from the oil. Phospholipase B is also useful for  
 CC preparing a dough or a baked product having improved property such  
 CC as increased strength, elasticity, stability and extensibility.  
 XX  
 SQ Sequence 464 AA;

Query Match 100.0%; Score 2501; DB 22; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 1e-215;  
 Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSTALLTGIGLLASGLASPVTSVTSREAPFGYKPGSKESIEMLKDKVENIWLILE 60  
 DB 1 mkstalltgigllasglaspsvtsvtsreapfgykpgskesienlkdkvenivwile 60  
 QY 61 NRSFDNLGGVRRQGLDNPINNGPFCNFKYKNSDPSGKGYCTQAKDYDSVFNDPQHSVTGN 120  
 DB 61 nrstdnlggvrrqgldnplngngpfcnyknsdpsgkgyctqakdydsvfndpshsvtgn 120  
 QY 121 NLRPYGTPNNCATASGVADQSGFLNAQLNDYPKLAPETATRVNGYITEEVEPTLV 180  
 DB 121 nlepygtpnngcatasgvadqsgflnaqlndypklapetatrvgngyiteeveptlv 180  
 QY 181 DLVDEFTTNSFCVPGTNPRLCALAGTAAGHGKNDDEFLNGYISKSIFEAAANEKG 240  
 DB 181 dlvdettntsfscvpgtpnprlcalagtaaghgknddflngyissksifeaanekeg 240  
 QY 241 VSWLNDGTNGFEFDSLEFTYVNOTSRNVVPEFQDAYLGVLPFESINPSCCGTN 300  
 DB 241 vswlndgtngfepdslefthyvnotsrnvvpvpefdaylglvlpfesinpsccgtn 300  
 QY 301 TNSMHPGTNWSVGEVFKQYDAIRGQPDWKTLLFTYDETGGFYDHPVPPPLAVRPDNL 360  
 DB 301 tnsmhpgtnwsygevfkqydairgqpdkllfitydetggfydhvppplavrpndl 360  
 QY 361 TYTETAKNGQKYLHFRDLGRMPTWVISPYSKGYIEQYDTPVTGKPAYSATSVLKT 420  
 DB 361 tytetakngqkylhfrdlgrmptwvispyskgyieqydgtpvtgkpaysatsvltk 420  
 QY 421 LGLYLDIEDFTPRVAHSPSFDHLIGTTLREDAPTALKTPTFTSV 464  
 DB 421 lglwldiedftprvahspsfdhlgttlredaptalktptftsv 464

RESULT 2  
 AAR07445  
 ID AAR07445 standard; protein; 483 AA.  
 XX  
 AC AAR07445;  
 XX  
 DT 01-FEB-1991 (first entry)  
 XX  
 DE Secretory signal peptide and protein with repressible acid  
 DE phosphatase activity.  
 XX  
 KW Acid phosphatase; yeast.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 EH Key Location/Qualifiers  
 FT Peptide 1..21

FT Protein /label=Secretory signal peptide  
 FT 22..483  
 XX /label=Protein has repressible acid phosphatase activity  
 PN JP02234681-A.  
 XX  
 PD 17-SEP-1990.  
 XX  
 PF 09-MAR-1989; 89JP-0057234.  
 XX  
 PR 09-MAR-1989; 89JP-0057234.  
 XX  
 PA (IKEM ) IKEDA MOHANDO KK.  
 XX  
 DR WPI; 1990-325618/43.  
 DR N-PSDB; AAQ06266.  
 XX  
 PT DNA contg. base sequences with promoter activity - used for  
 PT expression and secretory prodn. of pharmacologically important  
 PT protein gene  
 XX  
 PS Claim 1; Page 553; lpp; Japanese.  
 XX  
 CC Protein has repressible acid phosphatase activity, and carries a  
 CC secretory signal region allowing it to be produced in high yield into  
 CC the extracellular environment.  
 XX  
 SQ Sequence 483 AA;  
 Query Match 5.1%; Score 126.5; DB 11; Length 483;  
 Best Local Similarity 23.0%; Pred. No. 0.011;  
 Matches 67; Conservative 47; Mismatches 112; Indels 65; Gaps 14;  
 QY 134 ATASGKVADQ-SGLNAQLN-----DYPKLAPEATFQVNGYITEEVEPTLVLDL 182  
 DB 48 aaasavahtrsdvkaavnrfyvfildntdykavede-----slawl 92  
 QY 183 VDEFTTNSFCVPGTNPRLCALAGTAAGHGKNDDEFLNGYISKSIFEAAANEKGYS 242  
 DB 93 akqgisltnyws-lthsepnylasvagdy--faldddrfismpsnvnivldltdknis 149  
 QY 243 W-----LNVDGTWG-----EFEPDSLEFTYVNOTSR-SNVVPEVPEFQD 280  
 DB 150 waeyqehipygfgfngynqetfandvyrkhnplilfdnvisdkrlaniksfeefnd 209  
 QY 281 AYLGVLPFESINPSCCGTNWSMHPGTNWSVGEVFKQYDAIRGQPDWKT-TLLFTY 339  
 DB 210 lknktlpgyafitpnm-----tndghds-nikvagdwsfqlpdllddyfmdkdlvltf 264  
 QY 340 D--ETGGFYDHPVPPPL--AVRPDNLTYTETAKNGQKYLHFRDLGRMPTW 386  
 DB 265 denetygiknkvfsillggvipdhkgtl----sddtfydhysqatveanw 311  
 RESULT 3  
 AAW11866  
 ID AAW11866 standard; Protein; 1041 AA.  
 XX  
 AC AAW11866;  
 XX  
 DT 18-APR-1997 (first entry)  
 XX  
 DE ATH1 gene product, Athlp.  
 XX  
 KW Yeast ATH1; Athlp; vacuolar acid trehalase; hydrolysis; trehalose;  
 KW baking; frozen dough; dehydrated yeast; brewing; ethanol fuel;  
 KW molasses; corn syrup; freezing; wine; fermentation.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN US5587290-A.  
 XX

PD 24-DEC-1996.  
 PF 26-JUN-1995; 95US-0494714.  
 XX 26-JUN-1995; 95US-0494714.  
 PR (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Destruelle M, Holzer H, Kilonsky D;  
 XX WPI: 1997-064793/06.  
 DR N-PSDB; AAT61372.  
 XX  
 PT Mutant yeast deficient in prodn. of vacuolar acid trehalase - has  
 PT improved stress tolerance, esp. for dehydration, freezing and  
 PT alcohol concn., useful in baking, fermentation, etc.  
 XX  
 PS Disclosure: Column 17-24; 17pp; English.  
 CC This sequence is encoded by the yeast Ath1 coding sequence and represents  
 CC Ath1 which is a vacuolar acid trehalase. Yeast mutants which are  
 CC deficient in the production of functional Ath1p have reduced ability  
 CC to hydrolyse trehalose. These mutants are useful in baking (partic. in  
 CC frozen dough and dehydrated yeast products), in brewing, as a source of  
 CC trehalose (used as a protectant in foods and pharmaceuticals) and for  
 CC production of ethanol fuel from molasses or corn syrup. The yeast  
 CC mutants have improved tolerance of dehydration and freezing. They can  
 CC tolerate higher levels of ethanol and can grow at high cell densities  
 CC over a range of fermentable sugar concentrations, and so can produce  
 CC drier wines and complete fermentation more quickly.  
 XX  
 SQ Sequence 1041 AA;

Query Match 4.5%; Score 111.5; DB 18; Length 1041;  
 Best Local Similarity 19.7%; Pred. NO. 0.84;  
 Matches 89; Conservative 64; Mismatches 140; Indels 159; Gaps 24;

QY 22 VTSEY---TSVREAPGKPKCKESTENLKRVENIVWLLENRS-----FDNLGG 70  
 DB 446 vdeyhndvamasfslynghegid--eylryttwplknagaqfaykykynslgl 503  
 QY 71 VRROGLDNP-----INNGPFCNYKNASDPSSGKYCTOAKDYDSVFNPDHDSVTGNLLEF 124  
 DB 504 yetnltddpdefanhnngaf---tnagiktlkwtadignhigevdpkweiskdi-- 558  
 QY 125 YGTYTPNNGAIAAGKVVADQSGFLNAQLNDYKPAPEATRQV--MGYVTEEVPTLVDL 182  
 DB 559 ---yipr---ssnitleysg-----mussveikqadvtlmvyipygindesi--lna 604  
 QY 183 VDEFTTFNSWFCV--PGTPNPNRLCALAGTAAGH-----K 217  
 DB 605 ikdiyysergaaspantypfvaaag-llnhgsasqsylykswlpylrapfaqfseq 663  
 QY 218 NDDDFLNYGISSKSI--FEANEK-----GVSWLNYDCTNGEPEPDSLFFTYVNOTSR 269  
 DB 664 sddnflngitgafpftangflgslifgitgiry---syevdpdt---kkinrlirf 717  
 QY 270 NVVPEVNFODAYLVLP-----KFSYINPSCCGTNTNSMHPGTGNVSYGEVFKQIYDA 323  
 DB 718 n-pie-----lplpggtairnfkymnp----- 739  
 QY 324 IROGPQWDXTKLLFYDETCGFYDH-----VPPPLAVRPDLNLTETAKNGOKYT 373  
 DB 740 -----vidliidhngtviwhksgdvpilkiprslidhqdinfyngsenerkpn 789  
 QY 374 LH---FDRLGGRRPTWVISPYSKGYIEQYT 402  
 DB 790 lerrdvdrvgdpm-----zmdrygt 809

RESULT 4  
 AAR38697

ID AAR38697 standard; Protein; 579 AA.  
 XX  
 AC AAR38697;  
 XX  
 DT 25-NOV-1993 (first entry)  
 XX  
 DE PSY875 swine parvovirus B gene product.  
 XX  
 KW Attenuated; vaccine; herpes virus; non-primate; live; safer; IBR;  
 KW infectious bovine rhinotracheitis; MDV; Marek's disease virus; fowl;  
 KW pavedo-rabies; swine.  
 XX  
 OS Swine parvovirus.  
 XX  
 PN US5223424-A.  
 XX  
 PD 29-JUN-1993.  
 XX  
 PF 27-JUL-1988; 88US-0225032.  
 XX  
 PR 06-SEP-1985; 85US-0773430.  
 PR 27-JAN-1986; 86US-0823102.  
 PR 17-JUL-1986; 86US-0887140.  
 PR 02-SEP-1986; 86US-0902887.  
 PR 20-NOV-1986; 86US-0933107.  
 PR 27-JUL-1987; 87US-0078519.  
 PR 27-JUL-1988; 88US-0225032.  
 XX  
 PA (PRUT-) PROTECH RES & DEV.  
 XI  
 PI Chiang CH, Cochran MD, Macdonald RD;  
 XX  
 DR WPI: 1993-219585/27.  
 DR N-PSDB; AAQ42754.  
 XX  
 PT Recombinant fusion proteins for vaccine - comprises antigenic  
 PT sequences fused to viral sequences e.g. pseudo-rabies virus, used  
 PT as vaccines  
 XX  
 PS Disclosure; Fig 11; 127pp; English.  
 XX  
 CC The sequence is that encoded by the swine parvovirus B gene  
 CC sequence from PSY875 which may be used in the prodn. of attenuated  
 CC non-primate herpes viruses. These can be used as live vaccines and  
 CC provide a safer vaccine than currently available for e.g. pseudorabies  
 CC virus of swine, infectious bovine rhinotracheitis (IBR) virus or Marek's  
 CC disease of fowl.  
 XX  
 SQ Sequence 579 AA;

Query Match 4.4%; Score 109.5; DB 14; Length 579;  
 Best Local Similarity 21.4%; Pred. NO. 0.51;  
 Matches 86; Conservative 44; Mismatches 151; Indels 121; Gaps 22;

QY 110 FNDPDHSHVGTGNL-----EFY-----GTYTP-----NNGAIAAGKVVADQ 145  
 DB 121 fndpdqqlismnteinlsvfegelfnvlkitesatpspskiynditasmvaldtn 180  
 QY 146 GFLNAQLNDYKPAPEATRQVMGYVTEEVPTLVDLVDEFTTFNSWFCVPGTNPRL 205  
 DB 181 ntl-----pytpaaprset---lgfy--pwlpt-----kptgyrylscirlnpp--- 221  
 QY 206 CALAGTAAGHGKNDDEFNLNYGISKSIFFAANEKGVSWLNYDG---TNG--EFEPDSLFF 260  
 DB 222 -----tytgsgsqidsigtglhshdinfytienavpilhltgdefstgyhfdtkplk 276  
 QY 261 TYVNOTSRNVVPEVNFQ-----DAYLVLPKFSYINPSCCGTNT--NSMHPGTGNVSGE 314  
 DB 277 thswqtnrslgipkpvltteptegdqhgptlp-----gantrkyghgtinnsyte 326  
 QY 315 VFVKQIYDAIROGP--QWOKTLLFITVDETCGFYDHVPPPLAVRPDLNLTETAKNGOKYT 373  
 DB 315 VFVKQIYDAIROGP--QWOKTLLFITVDETCGFYDHVPPPLAVRPDLNLTETAKNGOKYT 373



XX PS Examples; Page 16-23; 57pp; English.

CC The invention relates to an acid trehalase (ATH1) gene-disrupted haploid yeast produced by gene manipulation, where the diploid is a practical baker's yeast. The new yeast strain is useful in dough which, upon subsequent fermentation and baking after a period of freezing, produces a high quality bread or sweetened bun. The new yeasts have a reduced ability of trehalose degradation during fermentation and thus an increased intracellular osmotic pressure, providing resistance to long -term freeze-storage when used in low and high-sugar dough. The dough makes better bread than dough using prior art baker's yeast. The new strain and ATH1 non-disrupted gene strains were used to prepare dough, and the new strain showed resistance to freezing. The present sequence represents an ATH1 gene product.

XX SQ Sequence 1041 AA;

Query Match 4.4%; Score 109.5; DB 20; Length 1041;  
Best Local Similarity 19.5%; Pred. No. 1.3; Indels 159; Gaps 24;  
Matches 88; Conservative 65; Mismatches 140;

QY 22 VTSEY---TSVREAPFGYKPGSKESIEINKKVENIVMLLENRS-----FDNILGG 70  
DB 446 vdeyhindhvamasfsiyinhghegidd-eylryttwpliknaaafftaykynsslgi 503  
QY 71 VRRQGLDNP-----INNGPFCNYKNASDPSSGKYCTQAKDYDSVPNDPDSHVTGNLLEF 124  
DB 504 yetynltpdefahinnaf---tnagiktikwatdignhlgvevdpkwselskdi-- 558  
QY 125 YGYTPNNGATASGVVADQSGFLNAQNDYFKLAPPEATPQV--MGYYTEEEVPTLVLDL 182  
DB 559 ---yipr---sssnitleysg-----mussveikqadvtlmvyplg/indexi--lnna 604  
QY 183 VDEFTTFNSWFCV-PGFTNPRLCALACTAAGHG-----K 217  
DB 605 ikdlytysergsaapantpyvfvaaag-llnngsssqsylykswlpylrapafqseq 663  
QY 218 NDDDFLNYGISKST-----FEAANEKGVSWLNVDGTNGEPEFDSLFTYYVNOFSRS 269  
DB 664 sddnftngltapafitangfglqsnlfgitgiry---syevdpdt---kkinrlrf 717  
QY 270 NVVPVENFQDAYLGLVP-----KFSYINPSCCGTNTNSMHPGTGNVSGVEFVKQIVDA 323  
DB 718 n--ple-----lplpgglalnrfkymn-----qvld- 742  
QY 324 IRQGPQWDKTLFFTYDGTGTFYDH-----VPPPLAVRPNLTYTETAKNGQKYT 373  
DB 743-----iiidhngtiwhksgdvpshikipnrslihdqinfngsenerkpn 789  
QY 374 LH---FDRLGGRMPWTWISPSYKKGIEQYGT 402  
DB 790 lerrdvdrvgdpm-----rmdrygt 809

RESULT 7  
AAG30820  
ID AAG30820 standard; Protein; 626 AA.  
XX AC AAG30820;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 36914.  
XX KW protein identification; signal transduction pathway; metabolic pathway;  
XX KW protein identification assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.

PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149802.  
PR 23-AUG-1999; 99US-0149830.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159284.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.3%; Score 108.5; DB 21; Length 626;  
Best Local Similarity 21.3%; Pred. No. 0.7;

Matches 99; Conservative 55; Mismatches 183; Indels 127; Gaps 23;

QY 38 PGSKESIEHLKDKVENIVWLILENRSFONILGVRQGLNDPNNPGFCNTK---NASDP 94  
Db 102 pdlgaslekhddvk---fsvydqkftn-ygnaraggads-----fknyskdgnvtd 151  
QY 95 SSGKTCQAKDYDSVFN--DpDHSVTGNLEFYCTYTPNNGALASGVVADQSGFLNAQL 152  
Db 152 sfrrysrnaaghdhdkftvygensnvveegfnsy9tf-----gtggagdfnynnnnpts 207  
QY 153 NDYPKLAPPEATRQVMGYTTEEVPTLVDLDEFTTFNSWFCVPGTPNPNRLCALAGTA 212  
Db 208 rfcaysdsgngrsqtktythe-----anang 235  
QY 213 ---AGHGN---DDDFLNGISSKSIPEANEGVSWLNDGTNGEPEPDSLEFFIVNQ 265  
Db 236 qstfygkngngvpeftsygvsnnvi-----gsfisy-gesgnaand-----tftsy 283  
QY 266 TSSNNVVPVENFFQDAYLG--VLPKF-SVINPSCCGTNTNSMHPGTN-----VSXGEV 315  
Db 284 gsdgn-vpqqnnfnnygasnaavdtfanyrdkanvgddsfsayakdsenkvfnvyggs 342  
QY 316 F-VQIYDAIRQCPQWDKTLFIYDETFGGFYHPPPLAVRPNLNYTTAKNGOKYT 373  
Db 343 fupgsetfgygkaegsk-lsfktytpnstfdkayakkvafakynvattantvcdgkt 401  
QY 374 LH-----EDRLGGR--MPTWIS--PY--SKGVIQYG- 401  
Db 402 vnkwpqkffresslkegtvimpdkdkmpkrslfprsiitkpfstsklgeikrith 461  
QY 402 -----TDPVTCKPAPYSATSVLTKTLGLWDIEDFTPRV 434

Db 462 avenstmgiltdavteceppsygetkrcvgsaedmidatsv 505

RESULT 8

AAV58162

ID AAV58162 standard; Protein: 532 AA.

AC AAV58162;

XX 07-MAR-2000 (first entry)

DT Adeno associated virus AAV5 capsid protein VP3.

DE

XX Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;

XX promoter; Rep protein; capsid protein; regulation; transcription;

KW replication; chromosomal integration; tissue tropism; cellular receptor;

KW gene therapy; neutralising antibody; erythroid progenitor cell;

KW transduction; cancer; genetic disease; VP3.

XX

OS Adeno associated virus type 5.

XX

XX W09961601-A2.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US11958.

XX 28-MAY-1998; 98US-0087029.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA Chlorini JA, Kotin RM;

PI WPI: 2000-062707/05.

DR N-PSDB; AA249214.

XX Adeno-associated virus 5 based vectors and particles, useful for gene therapy.

XX Claim 27; Page 80-82; 9lpp; English.

XX This sequence represents the VP3 capsid protein of

CC adeno associated virus type 5 (AAV5). The invention relates

CC to vectors comprising a pair of AAV5 inverted terminal repeats

CC (ITRs) with a promoter between the ITRs. The vector may comprise

CC the viral genome, or subregions thereof, including sequences

CC encoding Rep proteins and capsid proteins, and is encapsidated

CC in an AAV5 particle. The non-structural Rep proteins Rep40

CC (AAV58163), Rep52 (AAV58168), Rep68 (AAV58164) and Rep78 (AAV58159) are

CC involved in regulation of replication and transcription, in addition to

CC the production of progeny genomes. Rep68 and Rep78 are also associated

CC with the stable integration of the viral genome into human chromosomes.

CC The three types of capsid protein VP1 (AAV58160), VP2 (AAV58161) and VP3

CC (AAV58162) assemble to form an icosahedral capsid, and differ from each

CC other by the use of alternative splicing and an unusual translation

CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2

CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are

CC likely to utilise distinct cellular receptors and are serologically

CC distinct. In a gene therapy application, therefore, AAV5 would allow for

CC transduction of a patient who already possess neutralising antibodies

CC either as a result of natural immunological defence or from prior

CC exposure to AAV2 vectors. The vectors may be useful for transducing

CC erythroid progenitor cells or cells lacking heparin sulphate

CC proteoglycans, which is very inefficient with AAV2-based vectors.

CC The vectors may also be useful for transducing cells with a nucleic

CC acid of interest in order to produce cell lines that could be used to

CC screen for agents that interact with the gene product of the nucleic

CC acid of interest. In addition to transduction of other cell types,

CC transduction of erythroid cells would be useful for the treatment of

CC cancer and genetic diseases which can be corrected by bone marrow

CC transplants using matched donors.

XX

SQ Sequence 532 AA;

Query Match 4.3%; Score 107; DB 21; Length 532;

Best Local Similarity 19.1%; Pred. No. 0.74;

Matches 66; Conservative 35; Mismatches 113; Indels 132; Gaps 16;

Qy 118 TGNLLE-----FYCTYTPNCAIASGVYADQSGFLNAQLNDYKPLAPEATQVM 168

Db 206 tgnmfeytynfeevphsfapsqnlflklaplvdgylrfvsttm---tgvgqfknla 262

Qy 169 GYITEEVPTLVLDLVEFTTFNSFCVPGPTNRLCALACTAAGHGKNDDELFLYGIS 228

Db 263 gya-----ntykntf---pgpmgrtq-----gwnlgsqvn-----a 292

Qy 229 SKSIFEANEKGVSWLNYDGTNGEPEPDSLEFFTVVQTSRNVNVEFF---QAYILG 284

Db 293 svsatfntnr-----mlelasyqvpqngmt---nnlgsntyalentmifnsqpanp 345

Qy 285 VLPKFSYINPSCCGTNTNSMHPNGVSY-----GEV 315

Db 346 tta--tylegnmlitsetqpvrnvayvggmatngssttapatgtynlgeivpgsv 403

Qy 316 FVKQIYDAIROGPQPDWKTLLFIYDETGFY-----DHVPPPLAVR----PDNL 360

Db 404 wmer--dvylgpiwak-----ipetgahfhpapamggfglkhhppmmllkntpvgni 455

Qy 361 TYTETAKNGOKYTLHFDRLGRRMPTWVISPSYKKGYYEYGTGDPVT 406

Db 456 t-----sfsdvpssfitgystgvt 476

RESULT 9

AAV58161

ID AAV58161 standard; Protein: 588 AA.

XX AAV58161;

XX 07-MAR-2000 (first entry)

XX Adeno associated virus AAV5 capsid protein VP2.

XX Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;

KW promoter; Rep protein; capsid protein; regulation; transcription;

KW replication; chromosomal integration; tissue tropism; cellular receptor;

KW gene therapy; neutralising antibody; erythroid progenitor cell;

XX transduction; cancer; genetic disease; VP2.

XX Adeno associated virus type 5.

XX

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "The codon encoding this residue is specifically

FT mentioned in the specification as an unusual

FT transaltion initiation codon"

XX W09961601-A2.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US11958.

XX 28-MAY-1998; 98US-0087029.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Chlorini JA, Kotin RM;

XX WPI: 2000-062707/05.

DR N-PSDB; AA249213.

XX Adeno-associated virus 5 based vectors and particles, useful for gene therapy.

PT

Claim 25: Page 79-80; 9lpp; English.

This sequence represents the VP2 capsid protein of adeno associated virus type 5 (AAV5). The invention relates to vectors comprising a pair of AAV5 inverted terminal repeats (ITRs) with a promoter between the ITRs. The vector may comprise the viral genome, or subregions thereof, including sequences encoding Rep proteins and capsid proteins, and is encapsidated in an AAV5 particle. The non-structural Rep proteins Rep40 (AAV58613), Rep52 (AAV58168), Rep68 (AAV58164) and Rep78 (AAV58159) are involved in regulation of replication and transcription, in addition to the production of progeny genomes. Rep68 and Rep78 are also associated with the stable integration of the viral genome into human chromosomes. The three types of capsid protein VP1 (AAV58160), VP2 (AAV58161) and VP3 (AAV58162) assemble to form an icosahedral capsid, and differ from each other by the use of alternative splicing and an unusual translation initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2 capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are likely to utilise distinct cellular receptors and are serologically distinct. In a gene therapy application, therefore, AAV5 would allow for transduction of a patient who already possesses neutralising antibodies either as a result of natural immunological defence or from prior exposure to AAV2 vectors. The vectors may be useful for transducing erythroid progenitor cells or cells lacking heparin sulphate proteoglycans, which is very inefficient with AAV2-based vectors. The vectors may also be useful for transducing cells with a nucleic acid of interest in order to produce cell lines that could be used to screen for agents that interact with the gene product of the nucleic acid of interest. In addition to transduction of other cell types, transduction of erythroid cells would be useful for the treatment of cancer and genetic diseases which can be corrected by bone marrow transplants using matched donors.

Sequence 588 AA:

Query Match 4.3%; Score 107; DB 21; Length 588;  
Best Local Similarity 19.1%; Pred. No. 0.87;  
Matches 66; Conservative 35; Mismatches 113; Indels 132; Gaps 16;

QY	118	TGNLLE-----	FTGTTTPNNGAIGSKVADSGFNLQAQLNDYPKLAPAEATQVM	168
Db	262	tgnnfeytnfeevpfhsfapsqnlflklapldvlyrfvstnn-----	tgvgvqfnklnla	318
QY	169	GYYTEEEVPTLVDLEFTTFNSWSCVPGFTNRLCALAGTAGHGKNDODFLNYGIS	228	
Db	319	grya-----	nljknwf--pgpmgrtq-----gwnlgsygnr-----a	348
QY	229	SKSIFEAANEKGVSWLNDYDGETEPDLSLFTTVNGTSSRNWVVENFF-----	QDAYIG	284
Db	349	svsafattnr-----	melegasyvppqngmt--nnlgsnstyalentmifnsqbpng	401
QY	285	VLPKFSYIINPSCCCTNTNSMHPITGNVSY-----	GBV	315
Db	402	tta--tylegnmltsetetqpvnrvaenvggmatngssttapatytnlqeiavgsv	459	
QY	316	FVKQIYDAIRQGPQWDKTLFIYDETGGFY-----	DHVPPLAVR-----PDML	360
Db	460	wmer--dvyldgpIwak-----	ipetgahfnpaspangfglkhpppmmllkntpvpgni	511
QY	361	TYTETAKNGQKTYLHFDRLGGRMPTWVTSPIRSKGYTEQYGDTPV	406	
Db	512	t-----	sfsdvpssffitaygstqvt	532

RESULT	10
10	10

AA58160

AA136160  
ID AAY58160 standard; Protein; 724 AA.

0  
0  
H  
0  
0  
4  
4  
4

X  
X

AC AAY58160;  
XX  
DT 07-MAR-2000 (first entry)

XX	Adeno associated virus AAV5 capsid protein Vp1.
DE	
XX	
DE	
KW	Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;
KW	promoter; Rep protein; capsid protein; regulation; transcription;
KW	replication; chromosomal integration; tissue tropism; cellular receptor;
KW	gene therapy; neutralising antibody; erythroid progenitor cell;
KW	transduction; cancer; genetic disease; Vp1.
XX	
XX	Adeno associated virus type 5.
OS	
XX	
XX	W09961601-A2.
PN	
XX	02-DEC-1999.
PD	
XX	
XX	28-MAY-1999; 99WO-US11958.
PF	
XX	
XX	28-MAY-1998; 98US-0087029.
PR	
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA	
XX	
XX	Chiorini JA, Kotlin RM;
PI	
XX	WPI: 2000-062707/05.
DR	N-FSDS; AA49212.
DR	
XX	
XX	Adeno-associated virus 5 based vectors and particles, useful for gene
PT	therapy -
PT	
XX	Claim 23; Fig 5; 91pp; English.
PS	
XX	

Query Match 4.3%; Score 107; DB 21; Length 724;

QY 118 TGNLLE-----FYGYTPNNCAIASGKVVADSGFLNQLNDYPKLAPPEATRQVM 168  
||||| : : : | : : : || : | : : :  
Db 398 tqnnfeytnfeevpfhsafpsanlfklaplvdavlrvfstnnp---tgaavfkknla 454



us-09-426-072-2.rag

wed Dec 5 09:18:15 2001

Qy 169 GYHTEEVPTLVLDVDEFTTFNSWFCVPGTNPRLCALAGTAAAGCKNDKDDFLNNGIS 228  
Db 455 grya-----ntYknwF---pdpmgtrg-----gwnlgsgvnr-----a 484  
Qy 229 SKSIFEAAANEKGVSWLNDGTNGEPEPDSLFTTYVNOTSRNVVVENFF-----QDAYLG 284  
Db 485 svsafattnr-----melegasyqvpqpmgnt--nnlgssntyalentmifnsgpangp 537  
Qy 285 VLPKSYINPSCCGTNTNSMHTGNVSY-----GEV 315  
Db 538 tta--tylegmmltsetqpnrvaynvggmatngssttaopatynlqelvpvgs 595  
Qy 316 FVKQYDAIROGPOMDKTLFTYDETGFF-----DHVPPPLAVR-----PDNL 360  
Db 596 wmer--dvyllgpiwak-----ipetgahfhpamgfglkhppmmmlkntpvpgni 647  
Qy 361 TYETAKNGQYTLHEDRLGGRMPTWVSPYSKKGYEQYGDPTV 406  
Db 648 t-----sfsdvpvssfitqstgqvt 668  
RESULT 11  
ID AAR29028 standard; Protein; 1257 AA.  
AC AAR29028;  
DT 20-APR-1993 (first entry)  
DE Bacillus thuringiensis toxin protein 33F2.  
KW Toxin protein; ant.  
OS Bacillus thuringiensis strain PS33F2.  
PN W09220802-A.  
XX 26-NOV-1992.  
XX 22-MAY-1992; 92WO-US04316.  
XX 22-MAY-1991; 91US-0703977.  
PR 25-NOV-1991; 91US-0797645.  
PR 12-MAY-1992; 92EP-0304228.  
XX (MYCO ) MYCOGEN CORP.  
PI Kennedy MK, Meier H, Payne JM, Randall JB, Uick HJ;  
XX WPI; 1992-415780/50.  
DR N-PSDB; AAQ31411.  
XX Toxin proteins isolated from Bacillus thuringiensis - for controlling  
PT ants. e.g. fire, carpenter, argentine and pharaoh ants  
XX Disclosure; Page 45; 71pp; English.  
XX Bacillus thuringiensis toxin protein 33F2 is useful as a method of  
CC biological control of ants, e.g. fire ants, carpenter ants,  
CC argentine ants and pharaoh ants, as an alternative to chemical  
CC insecticides.  
XX Sequence 1257 AA;  
Query Match 4.38; Score 107; DB 13; Length 1257;  
Best Local Similarity 19.38; Pred.No. 2.8; Mismatches 166; Gaps 23;  
Matches 97; Conservative 73;  
Qy 6 LLTGLGLASLG-LASPVTSFVTSVREAPFGYKPGSKESIKENLKDKNVIVWLLLEN--- 61  
Db 74 visvltlpevgtvasaastlvsvflwplkfgdkpnaknifeelkpgleallqgditnygd 133

RESULT 12  
ID AAR28811 standard; Protein; 1257 AA.  
XX AAR28811;  
AC 30-MAR-1993 (first entry)  
DT BT toxin 33F2.  
DE nematode worms; nematocides; nematocidal toxin; agriculture; plants;  
KW crops; pests; CryV proteins.  
XX Bacillus Thuringiensis.  
XX W09219739-A.  
XX 12-NOV-1992.  
XX 01-MAY-1992; 92WO-US03624.  
XX 03-MAY-1991; 91US-0693018.  
PR 31-JAN-1992; 92US-0830050.  
PR 23-APR-1992; 92US-0871510.  
XX (MYCO ) MYCOGEN CORP.  
XX Focerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;  
PI WPI; 1992-398866/48.  
XX N-PSDB; AAQ30937.  
XX New genes and toxins against nematodes - obtd. from Bacillus  
PT Thuringiensis isolates with nematocidal activity  
XX Claim 1(b) ; Page 54; 77pp; English.  
XX This sequence represents the Bacillus thuringiensis delta-endotoxin  
CC 33F2 having nematocidal activity. It was decoded from the appropriate  
CC DNA. Intact cells expressing this toxin may be used in nematocidal

Qy 62 ----RSEFNIILGVRQGLDNPINNGPFCNYKNASDPSSGKYCTQAKDYDSVEND----- 112  
Db 134 ainqkfkdsiqktinlytv--aldnddyvtak-----tqlenlnsilttsdisif 180  
Qy 113 -PDHSYTGNNLEEGYTYTPNNGAIASGVV-ADQSGFLNAQLNDYPKLAPEEATRQVMGY 170  
Db 181 ipegyetg-gpyyamvanahllilirdaivnaeklgfskevdtbhky-----ikmtihn 234  
Qy 171 YTEEVPTLVLDVDEFTTF--NSWFCVPGTNPRLCALAGTAAAGCKNDKDDFLNNGIS 228  
Db 235 hteavikaflngldkfsidvnsy-----nkkany--- 264  
Qy 229 SKSIFEAAANEKGVSWLNDGTNGEPEPDSLFTTYVN-----QTSRSNVVVENF 277  
Db 265 ikgmtemvldivalvptfdpdhyqkvelefrttispiyqvpkmmntsssiivpsdif 324  
Qy 278 FQDAYLGVLPK--FS-----YINPSCCGT-----NTNSMHTGN 309  
Db 325 h---yggdlvklefstrtdndglakiftgirtfykspnthethyhdvfyntqs---sgn 378  
Qy 310 VSYG-----EVFVKQIYD-----AIRQGPOMDKTLF 336  
Db 379 isrgssnpiidlnnpiistcrlnsfykaagssvlnfkdgtygafaqap----- 430  
Qy 337 ITYDETGFGYDH--VPPPLAVRPDNLTYETAKNG-----QKYTLHEDRLGGRMPTWVI 388  
Db 431 -----tggadwhsflesdgapeghkinylytspadtlrdfinvytlitstpeisteiki 485  
Qy 389 SPY-SKKGYIEQYGTDPVTGKP 409  
Db 486 kgfpaekgyikngimkygk 507

RESULT	14
AA	AA44203
ID	AA44203 standard; Protein; 1257 AA.
XX	
XX	AA44203;
XX	
XX	09-DEC-1995 (first entry)
DT	
DT	
XX	Bacillus thuringiensis crystal protein 33f2.
DE	
DE	
XX	
XX	
KW	Crystal protein; delta endotoxin; acaricide; pesticide;

QY	389	SPV-SKGYIEQGTDPVTGKP	409
	:	:     :	
DBD	486	kfpfpaekgyikngimkygkp	507
RESULT 15			
XX	AAW13886		
ID	AAW13886	standard; protein;	1257 AA.
XX	AAW13886;		
XX	14-MAY-1997	(first entry)	
DT	XX		
XX	33F2 toxin.		
DE	XX	Toxin; ant;	Bacillus thuringiensis; hymenopteran pest; pharaoh ant;
XX	XX	biological control;	Monomorium pharaonis; delta-endotoxin; lepidoptera;
KW	insect.		
KW	XX		
OS	Bacillus thuringiensis isolate PS33F2.		
XX	US5596071-A.		
PN	XX		
XX	21-JAN-1997.		
PD	XX		
XX	22-MAY-1991;	91US-0703977.	
PF	XX		
XX	24-NOV-1993;	93US-0158232.	
PR	22-MAY-1991;	91US-0703977.	
PR	25-NOV-1991;	91US-0797645.	
PR	22-MAY-1992;	92US-0887980.	
XX	(MYCO ) MYCOGEN CORP.		
PA	Foncerrrada L;	Fu J,	Kennedy MK,
XX		Meier H,	Payne JM;
P1		Schwab CE	Ulick HT;

```
DR WPI: 1997-107615/10.
DR N-PSDB; AAT60071.
XX
PT Bacillus thuringiensis toxin - active against hymenopter pests
XX
PS Disclosure; Column 57-64; 64pp; English.
XX
CC This sequence represents the 3f2 toxin isolated from the Bacillus
CC thuringiensis (B.t.) isolate PS3f2. B.t. is a gram-positive, spore
CC forming, soil bacterium, characterised by paraspinal crystalline protein
CC inclusions. These proteins can be highly toxic to pests, and have been
CC used to produce insect resistant plants. The previously isolated B.t.
CC delta-endotoxins were mainly active against lepidopteran insects,
CC however the proteins of the invention are active against hymenopter
CC insects. This sequence is an example of a toxin of the invention, for
CC which the sequences shown in AAM13888 and AAM13871 represent the generic
CC formulae. As the toxins of the invention are active against hymenopter
CC pests, they can be used for the biological control of ants, particularly
CC pharaoh ants (Monomorium pharaonis).
XX Sequence 1257 AA;
QT
Query Match 4.3%; Score 107; DB 18; Length 1257;
Best Local Similarity 19.3%; Pred. No. 2.8; 166; Indels 166; Gaps 23
Matches 97; Conservative 73; Mismatches 73;
QY 6 LITGLIGLLASLG-LASPVTSEYTSVREAPFGYKPGSKESTENIKDKVENIVHILN--- 61
Db 74 vlsvltlpevgvtvasastivsfwpkfgdkpnaknifeeklpqlealqqdntynqd l33
QY 62 ----RSDNTLIGVRROQLDNPINNGPFCNYKNASDPSSGKYCTQAKDYDSVFND----- 112
Db 134 aingkkfdstiqtkinlytv--aidndnyvtrak-----tqlenlnsltsdisif 180
```

Db 181 ipeygetg-glpyyamvanahilllrrdaivnaekligfsdkevdtchky-----ikmtihn 234  
QY 171 YTEEEVPTLVLDLVDFTTF--NSWFSCVPGPTNPNRLCALAGTAAGHGKNDLDFLNYGIS 228  
Db 235 hteavikafingldkfksldvnsy-----nkkany--- 264  
QY 229 SKSIFEAAANEKGVSWLNYDGTNGEFPDLSLFTYVN-----QTSRSNVVPPVENF 277  
Db 265 lkgntemvidlvalwptfdpdhygkeveleftrtisspiyqvpvknmqntsssiivpsdlf 324  
QY 278 FQDAYLGVLPK--ES-----YINPSCCGT-----NTNSMHPGTGN 309  
Db 325 h---yqgdvlvklefstrtdndglakiftgirntfyskpnthetyhvdfsyntqs---sgn 378  
QY 310 VSYG-----EVFVKQIYD-----AIROGQPQWDKTLF 336  
Db 379 lsrqssnpipldlnnpiilstciirnsfykaigssvlvnfkdgtggyafaqap----- 430  
QY 337 ITYDETGGFYDH--VPPPLAVRPDNLTYTETAKNG-----QKYTLHFDRGLGGRMPTWVI 388  
Db 431 -----tggawdhsfiesdgapeghklniytspgdtdlrdfinvytlistptineisteckl 485  
QY 389 SPY-SKKGIEOYGTDPVTGKP 409  
Db 486 kgfpaekgyiknqgimkygkp 507

Search completed: December 4, 2001, 15:07:35  
Job time: 36 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:07:00 ; Search time 19.47 Seconds  
(without alignments)  
1815.357 Million cell updates/sec

Title: US-09-426-072-2

Perfect score: 2501  
Sequence: 1 MKSTALLGGLASGLAS.....GTTLRDAPIALKTPHTFSV 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453.5	18.1	533	2 E86207	hypothetical prote
2	436	17.4	462	2 T46205	hypothetical prote
3	435.5	17.4	514	2 T02648	probable phospholi
4	324.5	13.0	508	2 F70662	probable plcC prot
5	288	11.5	512	2 G70662	probable plcB prot
6	287.5	11.5	730	2 A26391	phospholipase C (E
7	287.5	11.5	730	2 B83540	hemolytic phosphol
8	286	11.4	512	2 H70662	probable plcA prot
9	267	10.7	692	2 A36143	phospholipase C (E
10	264	10.6	692	2 E83230	non-hemolytic phos
11	202.5	8.1	501	2 T35009	probable phospholi
12	167.5	6.7	422	2 B31776	hypothetical prote
13	158	6.3	280	2 J70987	probable plcD prot
14	130.5	5.2	483	2 J72179	acid phosphatase (
15	128	5.1	221	2 T35551	probable phospholi
16	119.5	4.8	417	2 J23389	acid phosphatase (
17	116	4.6	412	2 JN0319	acid phosphatase (
18	116	4.6	2178	2 S55805	alpha-toxin - Clos
19	112.5	4.5	488	2 A32235	poly(3-hydroxybuty
20	111	4.4	428	2 S32228	glutamate--ammonia
21	111	4.4	622	2 G86371	hypothetical prote
22	111	4.4	629	2 T07426	probable polygalac
23	110.5	4.4	729	1 A60006	coat protein vpl -
24	109	4.4	299	2 E70842	probable acid phos
25	109	4.4	1700	2 G86131	probable invasiv 2
26	108.5	4.3	626	2 T01485	probable polygalac
27	106.5	4.3	1341	2 S50366	probable membrane
28	105.5	4.2	962	2 C71617	SERA antigen/papai
29	105.5	4.2	1211	2 S54500	alpha.alpha-trehal

30	105	4.2	305	2 D83849	hypothetical prote
31	105	4.2	2229	2 T16199	hypothetical prote
32	104.5	4.2	729	1 VCPVNA	coat protein vpl -
33	104.5	4.2	991	2 T40229	arginyl endopeptid
34	104.5	4.2	1122	2 T18346	MG1 protein precu
35	104.5	4.2	1180	2 E86719	hypothetical prote
36	104.5	4.2	1704	2 A55426	gingipain R (EC 3.
37	103.5	4.1	1720	2 T43327	glucanate transpor
38	103.5	4.1	1181	2 C86349	hypothetical prote
39	103.5	4.1	1290	2 A57190	ehnerin precursor
40	102	4.1	578	2 D82902	hypothetical prote
41	101.5	4.1	362	2 H69785	mannan endo-1,4-be
42	101.5	4.1	559	2 T25418	hypothetical prote
43	101.5	4.1	800	2 S37387	internalin A precu
44	101.5	4.1	1203	2 E86266	hypothetical prote
45	101.5	4.1	1526	2 S49763	gingipain R (EC 3.

## ALIGNMENTS

### RESULT 1

E86207

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: E86207

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Cono, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: E86207

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-533 <STO>

A:Cross-references: GB:AE005172; NID:g8954023; PIDN:AAF82197.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 18.1%; Score 453.5; DB 2; Length 533;  
Best Local Similarity 30.9%; Pred. No. 2e-25;  
Matches 139; Conservative 77; Mismatches 167; Indels 67; Gaps 19;

QY	34	FGYKPGKESIE-----NLKDKVENIWLLENRFDNHLGGVR--RQGLDHPINNG	83
Db	13	FCYLLISSOSIEFKNSQRPKHIOGPIKTVVVVNNRNSFDHILGWLKSTRPEIDG--ITG	70
QY	84	PFCNKNADSPSGKYCTQAKDVSF--NDPDHVTGNLLEFYGYTPNNGAIAAGKVV	141
Db	71	KESNPLNVSDPNKSKKIFVSD---DAVFDMDFGHSFQAIREQIFG-----SNDTSGDPKRW	122
QY	142	ADQSGFLNAQLNDYFKLAPENATRVNGYTYEEVPTLVLDVDEFTFNSEWFCVPGPTN	201
Db	123	---NGFAQQSEMEFGMA-----KNVNSGKFEVLPVYTELANEGFVDFRWASVPTSTQ	174
QY	202	PNRLCALAGTAAGHGKNDDEFLNYGISKSIPEAAANEKGVSWLNYDGTNGEPEPSLFTT	261
Db	175	PNREYVHSATSHCCSNVKKDLVKGFPQKTFEDSLDENGSLF---GIYYQNIPTATFEFK	230
QY	262	YVNTQSRNVVPEV-----EFQDAYLGLVLPKFSYINPSCCGTN---TNSMHTPTGNVSYGE	314
Db	231	SLRL--KHULRFHSHVAKFKLDKAKGLKLPNSVYVQRYFDIDLFPANDDHFSDVDAQG	288
QY	315	VFWKQIYDAIROGPOQMDKTLFLITYDETGGFYDHPVPPPL--AVRPDNLTYTETAKNGQY	372

Db 289 RTVKVETLRSPPQWEMALLITYDEHGGFYDHPVTPVKGVNPDGII-----GDPDF 342  
QY 373 TLHFDRLGGRTWISPSKGYI---EQYGDTPDTGKPAPYSA---TSVLKTLGLYLDWI 427  
Db 343 YGFDRLGVRVPTFLISSWIEKGVIHEPEG-----PTPHSOFHSSIPATVKKLFNL 395  
QY 428 ED--FTPRVAHSPFDHL--IGTTLREDAP 453  
Db 396 KSHFLTKRDWAGTFEKYFRIDRSPQDCP 425  
RESULT 2  
T02605  
hypothetical protein T8P19.120 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T02605  
R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; S  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: 223008  
A:Accession: T02605  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-462 <CHO>  
A:Cross-references: EMBL:AL133315  
A:Experimental source: cultivar Columbia; BAC clone T8P19  
C:Genetics:  
A:Map position: 3  
A:Introns: 139/2; 307/1  
A:Note: T8P19.120

Query Match 17.4%; Score 436; DB 2; Length 462;  
Best Local Similarity 32.0%; Pred. No. 3.le-24;  
Matches 125; Conservative 61; Mismatches 165; Indels 40; Gaps 13;  
QY 78 NFINN---GPFCKYKASDPSSGKYCTQAKDYSDVNDPDHVSHTGNNLEFYGYTPNNGA 134  
Db 6 NPTNGVTGECN----PVPNSTQITCFTSDAEFVDPDGHSGFEAVQVFGS---GPGQ 58  
QY 135 TSKKVVADSGFLNALQNDYKLAPEEAATQVMGYTTEEVPTLVLDVDEFTTNSWFS 194  
Db 59 IPS-----NMGEVEQALS-----MPGNLSETVMKGFPEAPVYAEVLKFEAFVDRWFS 107  
QY 195 CVPGPTPNRLCALAGTAGHGKNDDEFNLNGISSKSIFFAANKSGVSLNVDGNTGFE 254  
Db 108 SIPGPTPNRLFYVSATSHGSTSHVKKQLAQGYPKITFDSLHSDIDF----GIYFQNI 163  
QY 255 PDSLEFFVYVQTSR-SNVVPVE-NFODAYLVLPKFSYINP---SCCGTNTNSMHTGN 309  
Db 164 PTFYRLNQLKYLIFNLHQYDLKFKDAKGLPSLVIEPRYFDLGLPANDDHPSHD 223  
QY 310 VSYGEVFVKIYDAIROGPQMDKTLFTITYDETGFGYDHPVPLAVRPDNLTYTTAKNG 369  
Db 224 VANGOKLVKEVYEAALRSSPQNNETLLVITYDEHGGFYDHPVTPVGPDPDGTGAPGF 283  
QY 370 QKYLTLHFDRLGGRTWISPSKGYIYQYDTPYTKPAPYSATSVLTKGLYLDWIED 429  
Db 284 FK---FDRUGVKTPTMWSPIQKGVVSEAKGT---ESSEYEHSSIPATIKKLFNLSS 337  
QY 430 --FTPRVAHSPFDHLIG--TTLREDAPAL 456  
Db 338 NFLTHRDAAWATFEDVWSHLTTPRTDCPMTL 368

RESULT 3  
T02648  
probable phospholipase C [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F12C20.9  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
C:Accession: T02648; H84665  
R:Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; S

submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.  
A:Reference number: Z14685  
A:Accession: T02648  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-514 <ROU>  
A:Cross-references: EMBL:AC005168; NID:g3426033; PID:g3426039  
A:Experimental source: cultivar Columbia  
R:Lib, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: H84665  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <STO>  
A:Cross-references: GB:AE002093; NID:g3426039; PIDN:AC32238.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: F12C20.9; At2g26870  
A:Map position: 2  
A:Introns: 156/2; 280/3

Query Match 17.4%; Score 435.5; DB 2; Length 514;  
Best Local Similarity 29.8%; Pred. No. 4e-24;  
Matches 130; Conservative 79; Mismatches 162; Indels 65; Gaps 17;  
QY 51 VENIVWLILNRSFDNITLGGVRRQGLDNPIN--NGPFCNYKNASDPSSGKYCTQAKDYDS 108  
Db 27 IKTIIVVMENRSFDMGLGMKK--LNPEINGVDGSESNPVSVSDPSRKIKFGSGSH-Y 83  
QY 109 VFNDPDHVSHTGNNLEFYGYT-----TPNNGAISGVVADSGFLNALQNDYKLAPEE 162  
Db 84 VDDPGHSGFOATREGVFGSNDTMDPPMPANGVQQA--YSDPDSGNMSA----- 130  
QY 163 ATROVMGYTTEEVPTLVLDVDEFTTNSWFSVCPGPTPNRLCALAGTAGHGKNDDEF 222  
Db 131 ---SVNMGEPDVKPVYKSLVSEFAVDFRWFASVPSSTQPNRMFVHSGTSAGATSNPNIS 187  
QY 223 LNYGISKSIFFAANKSGVSLNVDGNTGFEFEDSLFFTYVNOT-----SRSNVVPVE 275  
Db 188 LAGYFQRTIFDNLDEEFSF---GIYQNPVAVFLVQSLRKLKYVFKFHSYGN---- 238  
QY 276 NFODAYLVLPKFSYINPSCCGT---NTNSMHTGNVSYGEVFVKIYDAIROGPQMDK 332  
Db 239 SFKDHAQKGLPAYTVEQRYMPTLLEPASDDHPSHDYVQOQKFEKVEYETLRASPOWNE 298  
QY 333 TLLFTYDETGFGYDHPVPLP--AVRPDNLTYTTAKNGQKYLTHFDRLGGRTWISP 390  
Db 299 TLLITYDEHGGFYDHPVTPVRNVPSPDGIV-----GDPFLFQPNRLGIRVPTIAVSP 352  
QY 391 YSKKGYIEQYDTPYTKGPAP---YSATSVLTKGLYLDWIED--FTPRVAHSPFDHL-- 443  
Db 353 WIEKGV-VHGNP---GSPFSEYEHSSIPATVKKLFNLSSPFLTKRDEWAGTFENILQ 408  
QY 444 IGTTLREDAPIALKTP 459  
Db 409 IRKEPRDTCPETLPEP 424

RESULT 4  
F70662  
probable plcC protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: F70662  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

100

---

Db 51 VQHVILMOENKSFHFGHLNGVR--GENDP-----RALKRODGRPVWY---Q 94  
QY 105 DYDSVNDPDHVSVTGNLFFGYTYTPNNGAIASGVVADQS-----GFLNAQLND--- 154  
Db 95 NYKFEFS-PYHWDT-----KVTSAQVSSQNHEWSAFHAINQGRNCKWM 138  
QY 155 ---YPKLAPAEATQVMGYTYTEEVPTLVLDLVEFTFNFSVCPGPTNPNRLCALAGT 211  
Db 139 AVQVP-----BAMGYFKRGDIPYYALADAFITLCEAYHQSMGMPTNPNRLYHMSGR 189  
QY 212 AAGHGKNDDELNYGISKSTF-----EAAKEKGVSWLNVDGTNGEFEPDSLEF 260  
Db 190 AAPSGDGKDVHIGNDMGCGTIGASGVTDWTTTYPRLSAGVDNRVYQ---EGGYRSSSLWY 247  
QY 261 TYVNOTSRNSVVPVNF-----FQDAYL-----GVLP 287  
Db 248 LYVDAYWKYRLOEQNNYCNALAWFRNFKNAPRDSDLWQRAMLAGVDOLRKDVQENTLP 307  
QY 288 KFSYINPSCCGTNTNSMHTGNSYGEVEVKQIYDAIROGPO-WDKTLLFTFYDETGGFY 346  
Db 308 QVSWIVAPYCYE---HPWMPGSEGYVTVRLEALTSNPEVWARTVILNYDEGDFY 363  
QY 347 DHVPPPLAVRPNLTYTETAKNGQKTYLHFR-----LGGRMPTWVISPYSKKGYIDQY 401  
Db 364 DHASAPVPPMKDVGVLSTVSTAGE---IEVSSGLPIGLGHRVPLIAISPWKGGKV--- 416  
QY 402 TDPVTGKPAPIYATSVLKTL-----GYLWDIEDF-----TPRV 434  
Db 417 -----SAEVFDHTSVLRLERFGLVEENISPWRAVCGDLTSLFDFOGAGDTQVAPDL 470  
QY 435 AHSFSDHLTGTLTRED 452  
Db 471 TNVPQSD-----ARKEDA 483

## RESULT 7

B83540  
hemolytic phospholipase C precursor PA0844 [imported] - Pseudomonas aeruginosa (strain H  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83540  
Rstover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: B83540  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-730 <STO>  
A:Cross-references: GB:AE004519; GB:AE0046736; PIDN:AAG04233.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: pLCH; PA0844

Query Match 11.58; Score 287.5; DB 2; Length 730;  
Best Local Similarity 23.9%; Pred. No. 4.3e-13;  
Matches 119; Conservative 62; Mismatches 156; Indels 161; Gaps 22;

QY 51 VENTIVLILNENSFONILG---GVRRQGLDNPINNGPFCNYKNASDPSSGK---YCTQAK 104  
Db 51 VQHVILMOENKSFHFGHLNGVR--GENDP-----RALKRODGRPVWY---Q 94  
QY 105 DYDSVNDPDHVSVTGNLFFGYTYTPNNGAIASGVVADQS-----GFLNAQLND--- 154  
Db 95 NYKFEFS-PYHWDT-----KVTSAQVSSQNHEWSAFHAINQGRNCKWM 138  
QY 155 ---YPKLAPAEATQVMGYTYTEEVPTLVLDLVEFTFNFSVCPGPTNPNRLCALAGT 211  
Db 139 AVQVP-----BAMGYFKRGDIPYYALADAFITLCEAYHQSMGMPTNPNRLYHMSGR 189

QY 212 AAGHGKNDDELNYGISKSTF-----EAAKEKGVSWLNVDGTNGEFEPDSLEF 260  
Db 190 AAPSGDGKDVHIGNDMGCGTIGASGVTDWTTTYPRLSAGVDNRVYQ---EGGYRSSSLWY 247  
QY 261 TYVNOTSRNSVVPVNF-----FQDAYL-----GVLP 287  
Db 248 LYVDAYWKYRLOEQNNYCNALAWFRNFKNAPRDSDLWQRAMLAGVDOLRKDVQENTLP 307  
QY 288 KFSYINPSCCGTNTNSMHTGNSYGEVEVKQIYDAIROGPO-WDKTLLFTFYDETGGFY 346  
Db 308 QVSWIVAPYCYE---HPWMPGSEGYVTVRLEALTSNPEVWARTVILNYDEGDFY 363  
QY 347 DHVPPPLAVRPNLTYTETAKNGQKTYLHFR-----LGGRMPTWVISPYSKKGYIDQY 401  
Db 364 DHASAPVPPMKDVGVLSTVSTAGE---IEVSSGLPIGLGHRVPLIAISPWKGGKV--- 416  
QY 402 TDPVTGKPAPIYATSVLKTL-----GYLWDIEDF-----TPRV 434  
Db 417 -----SAEVFDHTSVLRLERFGLVEENISPWRAVCGDLTSLFDFOGAGDTQVAPDL 470  
QY 435 AHSFSDHLTGTLTRED 452  
Db 471 TNVPQSD-----ARKEDA 483

## RESULT 8

H70662  
probable pica protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70662  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holfroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70662  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-512 <COL>  
A:Cross-references: GB:283860; GB:ALL23456; NID:g3261681; PIDN:CAB06148.1; PID:g17812  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: pica

Query Match 11.4%; Score 286; DB 2; Length 512;  
Best Local Similarity 28.6%; Pred. No. 3.3e-13;  
Matches 128; Conservative 40; Mismatches 161; Indels 118; Gaps 20;

QY 7 LTGLGLASL-GLASPVTSYTSVREAPFGYKPGSKESIEENLKDVENIVWLILENRSFD 65  
Db 10 LTGAGAAFLMDWAAPVIERKAYGA---GPCPG-----HLTD-IEHIVLLMOENRSFD 57  
QY 66 NILGVRROGLDNPINNGPFCNYKNASDPSSGK---CTQAKDYDSV----- 109  
Db 58 HYFETLSS-----TNGFNAASPATQGMWNPMTQALDPAGVTIFRDLTTRGP 105  
QY 110 -----FNPDHVSVTGNLFFGYTYTPNNGAIASGVVADQSGLFMAQLNDYPKLAPEEA 163  
Db 106 FLDGECVNDPEHQWVGMHLAW-----NGGA-----ND--NWLPAQA 139  
QY 164 TRQ-----VMGYTTEEVPTLVLDLVEFTFNFSVCPGPTNPNRLCAL-----AG 210  
Db 140 TTRAGYVPLTWGYTTRQDIPHYLLADFTTICDGYHCSLLTGLPNRLYLSANIDPAG 199  
QY 211 TAAGHKNDDELNYGISKSTF-----EAAKEKGVSWLNVDGTN-GEFEPDSLEFTYVNTSR 268  
Db 200 TDGGPOLVPEGFLPLQIQFSRIMPENLEDAVGSWKVYQNKGLGRFNTIPISNGLVQAFR 259  
QY 269 SNVVPVEN-----FFQDAYLGVLKFKFSYINPSCCGTNTNSMHTGNSVSYGEV 315



us-09-426-072-2.rpr

Wed Dec 5 09:18:18 2001

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: E83230  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-692 <STO>  
A:Cross-references: GB:AE004754; GB:AE004946; PIDN:AAG06707.1; GSPDB:CN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: p1cn; PA3319

Query Match 10.6%; Score 264; DB 2; Length 692;  
Best Local Similarity 25.8%; Pred. No. 2.1e-11;  
Matches 108; Conservative 59; Mismatches 163; Indels 88; Gaps 19;

Qy 46 NLKDKVENIVLILNRSFDNILG---GVRROG--LDNPIINGPFCNYKNASDPSSCKY-99  
Db 44 NLKD-VEHVILMQENSRFDHYFGTLKGVGRGDRMAIPLDQGRVWHQKSGKEILLPYH 102  
Qy 100 ----CTQAKDYSVFNDDHSVTGNLFEYGYTPNNGATASCKYVADQSGFLNAQLNDY 155  
Db 103 FDTSTTSQRVDG-----TPHTWPD-----QAANNEGRMD--133  
Qy 156 PKLAPEEATQVMGYTEEVPTLDVDEFTFNSMFSVCPGPTNPNRLCALACT---A 212  
Db 134 -KWLPAK-TERSLGYKQDIAFOFAMANAFTICDAYHCSFGQGTNPNRLFLMTGTNDPL 191  
Qy 213 AGHG---KNDDF---LNVGISKSIPEANEKGVSLNVDCTNGCEPEPDSL--FTYV 263  
Db 192 GQHGPGVTTNDHSDNGPVEGYTWTTPERLOAGITWRVYQDMADNFSNPLIGFRQYR 251  
Qy 264 NOTSRNVV-----PVENFFQDAYLGVLPKFSVNPSCCGTNTNSMHP-TGNVSYGEV 315  
Db 252 AAAPDSPLVINGLSTWKLDAKRDVLANSLPQVSI-----VAPAKYSEHPGSPSPINGAE 307  
Qy 316 FVKQIYDAIROGPO-WDKTLFLTYDGTGFDYDHPVPPPLA--VRPNLTYTET----- 365  
Db 308 YTSWLDALANPEVWSKALLVWFDENGDFDHPVAPPASLNKDGTLRGKTTADATLE 367  
Qy 366 --AKNGQKYLHFDRLGGRMPTWISPKKYIEQYCTDPVTGCKPAPYSATSVLKT 421  
Db 368 WHTKGDIRYRNOPYGLGRVPMVYISPSKGGWNS-----QVFDHTSVIRFL 415

## RESULT 11

T35009  
probable phospholipase C - *Streptomyces coelicolor* (fragment)  
C:Species: *Streptomyces coelicolor*  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35009  
R;Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1999  
A:Reference number: Z21565  
A:Accession: T35009  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-501 <SEE>  
A:Cross-references: EMBL:AL079355; PIDN:CAB45565.1; GSPDB:GN00070; SCOREDB:SC4C6.01  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOREDB:SC4C6.01

Query Match 8.1%; Score 202.5; DB 2; Length 501;

Best Local Similarity 26.7%; Pred. No. 4.1e-07;  
Matches 63; Conservative 28; Mismatches 78; Indels 67; Gaps 8;

Qy 210 GTAAGHGKNDLDFLNYGISKSIPEANEKGVSLNLY-----DGTNCE 252

Db 260 QRAADPRSNLARYGIATPYCDFAADVRAANRPKVSMLVNI-----LQSEHPALPVAGAV 315  
Qy 316 FVKQIYDAIROGPO-WDKTLFLTYDGTGFDYDHPVPPPLA-----VRPNLTYTET 366  
Db 316 SWYATRLTLLSNPAWEKALIVSYDENGDFDHPVAPPASLNKDGTLRGKTTADATLE 374  
Qy 367 KNGQKYLHFDRLGGRMPTWISPKKY 393  
Db 375 SGGIRGPL---GLGRVPCIVISPSYR 398  
RESULT 9  
A36143  
phospholipase C (EC 3.1.4.3), nonhemolytic - *Pseudomonas aeruginosa*  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 30-Nov-1990 #sequence\_revision 30-Nov-1990 #text\_change 08-Oct-1999  
C:Accession: A36143  
R;Ostroff, R.M.; Vasil, A.L.; Vasil, M.L.  
J. Bacteriol. 172, 5915-5923, 1990  
A:Title: Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from *Pse*  
A:Reference number: A36143; MUID:91008968  
A:Accession: A36143  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-692 <OST>  
A:Cross-references: EMBL:M59304; NID:g151495; PIDN:AAA25968.1; PID:g151496  
C:Keywords: phosphoric diester hydrolase

Query Match 10.7%; Score 267; DB 2; Length 692;  
Best Local Similarity 25.2%; Pred. No. 1.3e-11;  
Matches 114; Conservative 64; Mismatches 181; Indels 94; Gaps 20;

Qy 11 GLLASLGLASPTSYTSVREAPFGYKQSGKESTENLKDKNVENIVLILNRSFDNILG- 69  
Db 15 GTNAVYATSMPLSSIQNALAIPAHRHG-----NLKD-VEHVILMQENSRFDHYFGT 67  
Qy 70 --GVRROG--LDNPIINGPFCNYKNASDPSSCKY-----CTQAKDYSVFNDDHSVTGN 120  
Db 68 LKGVGRGDRMAIPLDQGRVWHQKSGKEILLPYHFTSTTSQRVDG-----115  
Qy 121 NLEFYGYTPNNGATASCKYVADQSGFLNAQLNDYPKLAPEEATQVMGYTEEVPTLV 180  
Db 116 -----TPHTWPD-----QAANNEGRMD--KWLPAK-TERSLGYKQDIAFOF 156  
Qy 181 DLVDEFTFNSMFSVCPGPTNPNRLCALACT---AAGHG---KNDDF---LNVGISK 230  
Db 157 AMANAFTICDAYHCSFGQGTNPNRLFLMTGTNDPLGQHGPGVTTNDHSDNGPVEGYTWT 216  
Qy 231 SIFEANEKGVSLNVDCTNGCEPEPDSL--FTYVNOTSRNVV-----PVENFFQDA 281  
Db 217 TYPERLOAGITWRVYQDMADNFSNPLIGFRQYRAAAPDSPLVINGLSTWKLDAKRDV 276  
Qy 282 YLGVLPKFSVNPSCCGTNTNSMHP-TGNVSYGEVYFVKQIYDAIROGPO-WDKTLFLTY 339  
Db 277 LANSLPQVSI-----VAPAKYSEHPGSPSPINGAEYTSWLDALANPEVWSKALLVWF 332  
Qy 340 DETGFGFDYDHPVPPPLA--VRPNLTYTET-----AKNGQKYLHFDRLGGRMPTWVI 388  
Db 333 DENGGDFDHPVAPPASLNKDGTLRGKTTADATLEWHTKGDIRYRNOPYGLGARVPMYVI 392  
Qy 389 SPYSKKYIEQYCTDPVTGCKPAPYSATSVLKT 421  
Db 393 SPWSKGGWNS-----QVFDHTSVIRFL 415

## RESULT 10

E83230  
non-hemolytic phospholipase C precursor PA3319 [imported] - *Pseudomonas aeruginosa* (stra  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83230

C:Species: *Kluyveromyces marxianus* var. *lactis*  
C:Date: 21-May-1990 #sequence\_revision 20-Jun-2000  
C:Accession: B31776  
R:Chang, Y.D.; Dickson, R.C.  
J. Biol. Chem. 263, 16696-16703, 1988  
A:Title: Primary structure of the lactose permease gene from the yeast *Kluyveromyces fragilis*  
A:Reference number: A92683; MUID: 89034136  
A:Accession: B31776  
A:Molecule type: DNA  
A:Residues: 1-422 <CHIA>  
A:Cross-references: AB:X06997; NID:g2856; PID:CAA30054.1; PID:g1325996

QY 342 TGGFY DRVPPPLAVRPDLNLTETAKNGQKYL -HFDR-----LGGRMPTWVI 388  
 ||||| ||| | : : : : ||| ||| : |||  
 Db 115 NGGFEDHVVPATA-----PAGTPGEVTVPDIDQVPGSGGIRGPIGLGRFVPCFVI 165  
 QY 389 SPYSK 393  
 |||||  
 Db 166 SPYSK 170

QY 389 SPYSK 393  
 III:  
 Db 166 SPYSR 170

RESULT 14  
JG7170

acid phosphatase (EC 3.1.3.2) Pho610 - yeast (*Kluyveromyces marxianus*)  
N:Alternate names: Pho610 protein  
C:Species: *Kluyveromyces marxianus*

P: Peter J. Rasmussen  
 C: Date: 04-Mar-2000 #sequence\_revision 04-Mar-2000 #text\_change 11-May-2000  
 C: Accession: J01719  
 R: Yoda, K.; Ko, J. H.; Nagamatsu, T.; Lin, Y.; Katbara, C.; Kawada, T.; Tomishige, N.;

A: Reference number: JC7179; MUID:20169637

A:Accession: JC7179  
A:Molecule type: DNA  
A:Residues: 1-483 <YOD>

A:Cross-references: GB:E02615  
A:Experimental source: strain Y-610  
C:Comment: This enzyme, highly glycosylated, is a covalently-linked cell-wall protein

rganic compounds.  
C:Genetics;  
A:Gene: pho610

C; keywords: cell wall; endoplasmic reticulum; glycoprotein; phosphoric monoester hyd

Query Match 5.2%; Score 130.5; DB 2; Length 483;  
Best Local Similarity 23.4%; Pred. NO. 0.071;  
Matches 68; Conservative 47; Mismatches 111; Indels 65; Gaps 14;

134 AIASGKVADQ-SGFLNAQLN-----DYPKLAPEATRQVMGYTTEEVPTLVDL 182

48 AMSSAVIAHKRTSDVAGAAVNRKFIIFLENIDIDAADE\*-----SLAWL 92

183 VDEFTTNSWFCVPGPTPNRLCALACTAAGHGKNDODFLNGISSKSIFEAAAEKGV 242

[illegible]





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:07:00 ; Search time 13.39 Seconds  
(without alignments)  
1270.536 Million cell updates/sec

Title: US-09-426-072-2

Perfect score: 2501  
Sequence: 1 MKSTALLTGLGLASGLAS.....GTTLRDAPIALKTPHTFSV 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	324.5	13.0	517	1	PHLC_MYCTU
2	318	12.7	700	1	PHLN_BURFS
3	308	12.3	514	1	PHLD_MYCTU
4	288	11.5	521	1	PHLB_MYCTU
5	287.5	11.5	730	1	PHLC_PSEAE
6	286	11.4	520	1	PHLA_MYCTU
7	264	10.6	692	1	PHLN_PSEAE
8	167.5	6.7	421	1	PHOX_KLUOLA
9	119.5	4.8	417	1	PHOA_PENCH
10	116	4.6	412	1	PHOB_PENCH
11	112.5	4.5	488	1	PHB_ALCFA
12	111	4.4	428	1	GLNC_BRANA
13	110.5	4.4	729	1	COAL_PAVP9
14	109.5	4.4	729	1	COAL_PAVPK
15	107	4.3	1257	1	CCAA_BACTU
16	106.5	4.3	1341	1	YL78_YEAST
17	105.5	4.2	1211	1	ATHI_YEAST
18	105.5	4.2	4590	1	FATH_HUMAN
19	105	4.2	749	1	COA2_PAVPN
20	105	4.2	939	1	HEX_ADEP3
21	104.5	4.2	625	1	BGAL_LACSK
22	104.5	4.2	729	1	COAL_PAVPN
23	104.5	4.2	991	1	CPGL_PORGI
24	104.5	4.2	1122	1	ADPL_MYCGA
25	101.5	4.1	800	1	INLA_LISMO
26	101.5	4.1	1203	1	ALAB_ARATH
27	101	4.0	1146	1	KTXA_KLUOLA
28	100.5	4.0	736	1	CPG2_PORGI
29	99	4.0	434	1	GLN2_HORVU
30	99	4.0	1645	1	OMP8_RICTY
31	98	3.9	430	1	GLN2_PEA
32	97.5	3.9	458	1	YM51_YEAST
33	97	3.9	558	1	PAHA_CAEL

34 97 3.9 584 1 D29B\_ARATH  
35 97 3.9 600 1 L765\_ARATH  
36 96.5 3.9 775 1 VP4\_ROTFL  
37 96.5 3.9 1006 1 RAT1\_YEAST  
38 96.5 3.9 2148 1 VTI1\_AEDAE  
39 96.5 3.9 4447 1 PRSK\_BACSU  
40 96 3.8 532 1 GSI\_NEUCR  
41 96 3.8 659 1 AMIA\_STRPN  
42 96 3.8 681 1 YADD\_SCHPO  
43 96 3.8 2329 1 YS89\_CAEL  
44 95.5 3.8 976 1 AMY\_BUTFI  
45 95.5 3.8 1783 1 Y468\_MYCGE

Q06737 arabidopsis  
Q04980 arabidopsis  
P39034 feline rota  
Q02792 saccharomyc  
Q16927 aedes aegypt  
P40803 bacillus su  
P38678 neurospora  
P18701 streptococc  
Q09838 schizosacch  
Q09624 caenorhabdi  
P30269 butyrivibri  
Q49460 mycoplasma

#### ALIGNMENTS

RESULT 1  
PHLC\_MYCTU STANDARD; PRT; 517 AA.  
AC P95245;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE PHOSPHOLIPASE C 3 PRECURSOR (EC 3.1.4.3).  
GN PLCC OR RV2349C OR MT2414 OR MTCY98.18C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544 (1998).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
Bisai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = 1,2-  
DIACYLGLYCEROL + CHOLINE PHOSPHATE.  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL PHOSPHOLIPASE C FAMILY.  
CC -!- MISCELLANEOUS: POLYMORPHISM WAS DISCOVERED IN THE PHOSPHOLIPASE  
PLCA/B/C REGION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; 283860; CAB06146.1; ALT\_INIT.  
DR EMBL; AE007081; AAK46707.1; ALT\_INIT.  
DR TIGR; MT2414; -.



RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RT Nature 393:537-544(1998).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=M.bovis; STRAIN=BCG / PASTEUR;  
 RX MEDLINE=99255698; PubMed=10320585;  
 RA Gordon S.V., Brosch R., Billault A., Garnier T., Eiglmeier K.,  
 RA Cole S.T.;  
 RT "Identification of variable regions in the genomes of tubercle bacilli  
 RT using bacterial artificial chromosome arrays.";  
 RT Mol. Microbiol. 32:643-655(1999).  
 CC [-] CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = 1,2-  
 CC DIACYLGLYCEROL + CHOLINE PHOSPHATE.  
 CC [-] SIMILARITY: BELONGS TO THE BACTERIAL PHOSPHOLIPASE C FAMILY.  
 CC [-] MISCELLANEOUS: POLYMORPHISM WAS DISCOVERED IN THE PHOSPHOLIPASE  
 CC PLCD REGION.  
 CC [-] CAUTION: THE GENE CODING FOR THIS PROTEIN SEEMS TO BE DEFECTIVE IN  
 CC STRAIN H37RV WHERE IT IS INTERRUPTED BY THE INSERTION OF AN IS6110  
 CC ELEMENT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z95890; CAB09340.1; ALT\_SEQ.  
 DR EMBL; AF007040; AAK46072.1; -.  
 DR EMBL; Y18606; CAB44656.1; -.  
 DR TIGR; M1799; -.  
 DR TubercuList; rv1755c; -.  
 KW Hydrolase; Signal; Complete proteome.  
 FT SIGNAL 1 37 POTENTIAL.  
 FT CHAIN 38 514 PHOSPHOLIPASE C 4.  
 SQ SEQUENCE 514 AA; 55770 MW; 9D99A466D312BA30 CRC64;

Query Match 12.3%; Score 308; DB 1; Length 514;  
 Best Local Similarity 28.0%; Pred. No. 8.7e-15;  
 Matches 124; Conservative 45; Mismatches 156; Indels 118; Gaps 20;  
 QY 9 GLGLASLGLASPVTSYTSVREAPFYKPGSKESIKDKVNIWILLENRFDNIL 68  
 DB 22 GAGALMS--FAGPVI-----EKAYGAGPCS---GHLTD-IEHFVFFMQENRSEDIHF 67  
 QY 69 GGVRRQLDNPINNGPCFNKYNKASDPSSGKYCTQAKDYDSV-----F 110  
 DB 68 GTL--SGTDGFTNVPSPLPQKQGNP-----MTQALDGTGTPWYRFDTRGPFPLDGACV 119

QY 111 NDPDHSVTGNLLEFYCYTTPNNGAIAAGKVVVAOSGFLNAQLNDYPKLAPEEAETROVMGY 170  
 DB 120 NDPDHS-----WVAHIESWNGGV-----NDNWLPAQ-----AKTRSAAHPTVMGY 160  
 QY 171 YTEEEVPTLDVDEFTTNSFCVPGTPNPRICALAGTAAGCKKDDDFL-----223  
 DB 161 YTRQDIPHYLLADATFVCDRIFCSVLGPTLPNRLWLSATIDPDGQNGGPELQSTFPQ 220  
 QY 224 ----NYGSSKISFEANEKGVSWLNDGTNGPEPEPDSLFFTYVNOTS-----267  
 DB 221 VRRFGWRIMPONLSDA-----GVSWKVY--RNKTLGPISSVLTYGSLVTFKQSDAPRSDL 274  
 QY 268 -RSNVVP--VENFFQDAYLGLVKFYSINPSCCGCTNTNSMHPITGNVSYGEVFKQIYDAI 324  
 DB 275 VRFVAPSPASPAANDVLANRLPRVSWIPNV-----LESEHPVAPAAAGAFANILRL 330  
 QY 325 RGQPG-WKTLFTTYDEYGGYDHPPLAVRPDNLITYETAKNGOKYTL-HFDR-----378  
 DB 331 LANPAVWEKTAIIVSYDENGFGFDHVPATA-----PAGTGEYTVTPDIDQVPGS 381  
 QY 379 -----LGGRMPTWVISPYSK 393  
 DB 382 GGIRGPIGLGFRVPCFVISPYSR 404  
 RESULT 4  
 ID PHLB\_MYCTU STANDARD; PRT; 521 AA.  
 AC P95246; Q50561;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE PHOSPHOLIPASE C 2 PRECURSOR (EC 3.1.4.3)  
 GN PLCB OR MPCB OR RV2350C OR M2415 OR MTCY98.19C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=H37RV;  
 RX PubMed=8757862;  
 RA Johansen K.A., Gill R.E., Vasil M.L.;  
 RT "Biochemical and molecular analysis of phospholipase C and  
 RT phospholipase D activity in mycobacteria.";  
 RL Infect. Immun. 64:3259-3266(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC [-] FUNCTION: HYDROLYZES SPHINGOMYELIN IN ADDITION TO







CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L11868: AAB59164.1; ALT\_INIT.  
 CC EMBL; L11868: AAB59165.1; ALT\_SEQ.  
 DR EMBL; U49511: AAC18943.1; -  
 DR EMBL; Z83860: CAB06148.1; ALT\_INIT.  
 DR EMBL; AE007081: AAK46709.1; -  
 DR EMBL; M57952: AAA63288.1; ALT\_FRAME.  
 DR EMBL; S69737: AAC60465.2; ALT\_FRAME.  
 DR TIGR; MT2416: -  
 DR Tuberculist; RV23516: -  
 KW Hydrolyase, Signal, Membrane, Antigen, Complete proteome.  
 FT SIGNAL 1 37 POTENTIAL.  
 FT CHAIN 38 520 PHOSPHOLIPASE C 1.  
 FT CONFLICT 327 327 T -> N (IN REF. 5).  
 FT CONFLICT 369 370 GE -> AQ (IN REF. 5).  
 FT CONFLICT 406 407 RG -> PR (IN REF. 5).  
 FT CONFLICT 516 520 SGLCS -> AGCAAEISR (IN REF. 5).  
 SQ SEQUENCE 520 AA: 56138 MW: 5844EF0C064288A7 CRC64;

Query Match 11.48; Score 286; DB 1; Length 520;  
 Best Local Similarity 28.6%; Pred. No. 3.3e-13;  
 Matches 128; Conservative 40; Mismatches 161; Indels 118; Gaps 20;

QY 7 LTGGLLASL-GLASPVTSSEYTSVREAPFYKPKGSKESINLKDKNVENIWLILENSFD 65  
 DB 18 LTGAAAFLLMDMAAPVIEKAYGA-----GPCPG-----HLTD-IEHIVLLMQENRSD 65  
 QY 66 NLGGVRRQGLDNPINPGFPCNYKNASDPSSGKY-----CTQAKDYDSV----- 109  
 DB 66 HYFGTLSS-----TNGFNAASPAFQMGWNPMTQALDPAGVTPRLDTRGP 113  
 QY 110 -----FNDPDHSDVTGNLFEYTYTPNNGAIGASGVYVADQSGFLAQLNPKLAPSEA 163  
 DB 114 FLDGECVNDPEHGWCMHLW-----NGGA-----ND-NWLPAQA 147  
 QY 164 TRQ-----VMGYTEEEVPLVDLVDDETFNFSFVCPGPTNPNRLCAL-----AG 210  
 DB 148 TTRAGPVPLTMGYTRQDIPIHYLLADFTICDGHCSLTLGLPNRLWLISANIDPAG 207  
 QY 211 TAAGHGKDDDFLWNGISKSIF-EAANEKGVSWLNYDGTN-GEFEPDLSFTTYVNTSR 268  
 DB 208 TDGQPQLVEPGLPLQOFSWRIMPENLEDAGVSWKYQNGKLGFRINTPISNNGLVQAFR 267  
 QY 269 SNVYFVEN-----FFQDAYLGLPKFYSINPSCCGTNTNSMHPGTGNYSGEV 315  
 DB 268 QAADRNRLARYGIATPYGDFADVPAANLKPVSILVFN-----LQSEHPALPALGAV 323  
 QY 316 FVKQIYDAIRGPQ-WDKTLFTYDTEGGYDHPVPLA-----VRPNLTYTTETA 366  
 DB 324 SMVATLRLISNPAVWEKTALIVSYDENGDFDHTPTTAPPTGPFVTPNID-AVPG 382  
 QY 367 KNGOKYTLHFDRLGGRMPTWTSYPSK 393  
 DB 383 SGGIRGPL-----GLGFVFCVISPYSR 406

## RESULT 7

ID PHLN\_PSEAE STANDARD; PRT; 692 AA.  
 AC P15713;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NON-HEMOLYTIC PHOSPHOLIPASE C PRECURSOR (EC 3.1.4.3) (PLC-N)  
 DE (PHOSPHATIDYLCHOLINE CHOLINEPHOSPHODIOLASE) (PHOSPHATIDYLCHOLINE-  
 DE HYDROLYZING PHOSPHOLIPASE C) (PC-PLC).  
 GN PLCN OR PA3319.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.

OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91008968; PubMed=2120196;  
 RA Ostroff R.M., Vasil A.I., Vasil M.L.;  
 RT "Molecular comparison of a nonhemolytic and a hemolytic phospholipase  
 C from *Pseudomonas aeruginosa*.";  
 RL J. Bacteriol. 172:5915-5923(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 15692 / PAOI;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Phan X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: HYDROLYSES PHOSPHATIDYL SERINE AS WELL AS  
 CC PHOSPHATIDYLCHOLINE.  
 CC -1- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O -> 1,2-  
 CC DIACYLGLYCEROL + CHOLINE PHOSPHATE.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL PHOSPHOLIPASE C FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M59304: AAA25968.1; -  
 DR EMBL; M30223: AAA25969.1; -  
 DR EMBL; AE004754: AAG06707.1; -  
 DR PIR: A36143; A36143.  
 KW Hydrolyase, Signal, Complete proteome.  
 FT SIGNAL 1 32 POTENTIAL.  
 FT CHAIN 33 692 NON-HEMOLYTIC PHOSPHOLIPASE C.  
 FT CONFLICT 18 18 G -> A (IN REF. 1).  
 FT CONFLICT 385 385 P -> A (IN REF. 1).  
 FT CONFLICT 584 585 ER -> DG (IN REF. 1).  
 SQ SEQUENCE 692 AA: 77144 MW: 6B16C0BEEB25CAEF CRC64;

Query Match 10.68; Score 264; DB 1; Length 692;  
 Best Local Similarity 25.8%; Pred. No. 1.8e-11;  
 Matches 108; Conservative 59; Mismatches 163; Indels 88; Gaps 19;  
 QY 46 NLKDKVENIWLILENRSFDNILG---GVRQGG--LDNPNNGFPFCNYKNASDPSSGKY- 99  
 DB 44 NLKD-VEHVILLMQENRSFDHYFTLKGVRGFGDMAIPLDQGRVWHQKSGKILPYH 102  
 QY 100 ---CTQAKDYDSVNDPDHSDVTGNLFEYTYTPNNGAIGASGVYVADQSGFLAQLNDY 155  
 DB 103 FDTSTTSAGRVG-----THTWPA-----QQAANEGRMD-- 133  
 QY 156 PKLAPEATRQVMGYTYTEEVPTLVLDVDETFNFSFVCPGPTNPNRLCALAGT---A 212  
 DB 134 -KWLPAK-TERSLGYKQDIAFOFAMANAFTICDAYHCSFGQGTNPNRLFTWTGTNDPL 191  
 QY 213 AGHG---KNDDDF---LNYGISKSIFFAANEKGVSWLNYDGTNCFEPPDSL--FFTYV 263  
 DB 192 GQHGGPVTTNDHDSNGPVEQGYTWYTPERLQAAGITWRVYQDMANFSDNPLIGFQYR 251  
 QY 264 NQTSRSNVV-----PVENFFQDAYLGLVLPKFSYINPSCCGTNTNSMHP-TGNVSYGEV 315  
 DB 252 AAAPDSPLIVNGLSLTKDLKRDVLANSLPQVSWI-----VAPAKYSEHPGSSPINGAE 307  
 QY 316 FVKQIYDAIRGPQ-WDKTLFTYDTEGGYDHPVPLA--VRPNLTYTTET----- 365

```

Db 308 YTSWLDALTANPEWNSKALLVMFENDGFFDHVAPPAAPSLNKGTLRGKTKTADATLE 367
QY 366 --AKNGQKYLHFDRLGRMPTWISPYSKGKYLEQYGTDPVTKGKPPAPYSATSKYLKTL 421
Db 368 WHTKGDIRNQPIGLGRVPMYVISPMSKGGWNS-----QVFDHTSVIRFL 415

RESULT 8
ID PHOX_KLULA STANDARD; PRT: 421 AA.
AC P08540;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE POTENTIAL ACID PHOSPHATASE (EC 3.1.3.2).
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHS 2359 / IPO 1267 / NRRL Y-1140;
RX MEDLINE=89034156; PubMed=3053697;
RA Chang Y.-D., Dickson R.C.;
RT "Primary structure of the lactose permease gene from the yeast
RL Kluyveromyces lactis. Presence of an unusual transcript structure.";
RL J. Biol. Chem. 263:16696-16703(1988).
CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN
CC ALCOHOL + ORTHOPHOSPHATE.
CC -!- SIMILARITY: VERY STRONG, TO ASPERGILLUS NIGER AND PENICILLIUM
CC CHRYSOGENUM PHOA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06997; CAA30054.1; ALT_INIT.
DR PIR; B31776; B31776.
KW Hypothetical protein; Hydrolase.
FT ACT_SITE 228 228 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 421 AA; 47193 MW; 2F7E614C943824E6 CRC64;

Query Match 6.78; Score 167.5; DB 1; Length 421;
Best Local Similarity 21.0%; Pred. No. 7.5e-05;
Matches 89; Conservative 65; Mismatches 147; Indels 123; Gaps 23;

QY 122 LEFGTGTTPNNGAIAAS-----GKVVADQSG-----FLNAQLNDYKLAPEATRQV 167
Db 31 LRTYSTISPSLSEIASASATEVAEVSDVEGAERFFIFLENTDYDKAAGDES---- 86
QY 168 MCYTEEBEVTPLVDLVEFTTFNSWFCVPGPTNPNRLCALAGTAAGHGKNDLNGI 227
Db 87 LSWLAEOGI-----TLTNWY--ALTHPSEPNYLASVGGDY--FALDDORTISMP 132
QY 228 SSKSIFFAANEKGVSMNYD-----GTNG-----EPEPDSLFFTYVNTQ 266
Db 133 NVSNIVDLLOTKGISWAEQVSHSPYAGFCQGMNFSNQETVASYDYVRKHNPILLFDNVNND 192
QY 267 SR-SNVVPEVNFQDAYGLVLPKFSYINPSCCGTWTNSMHPGTVSGEYFVKYQIDAIR 325
Db 193 TRLANIKNFEDFNNDVENEKLPQVAFITPNM-----TNDGHDOT-TIQFAGKWSKDFLAPLL 247
QY 326 QCPQW-DKTLFFITYDET-----CG-----FYDHPV----- 350
Db 248 ENDYFMEDTLVLLTFDENETYGKIKNVFSLILGGLVDPDLKTKDDTFYDHYVSQLASVEA 307
QY 351 ----PPLAVRPDNLITETAKNGQKYT-LHFDRLGRMPTWISPYSKGKIYQYGTDPV 405

```

```

Db 308 NMDPLPHLGHGDANDVLEIVANATNTINVEVD-----TTIMINE-TYIGLNDYINIE-- 358
QY 406 TGKPAP-----YSATSVLTKLYLWDIEDTPRVAHSPSFDHLIGTLTREDAPIALKTP 459
Db 359 --LPAFNTAINRNGQPILOSIKETWE-DEYSKQVSES----YTTSTTTTTSADV--DA 409
QY 460 HTFS 463
Db 410 ETFS 413

RESULT 9
ID PHOA_ASPNG STANDARD; PRT: 417 AA.
AC P34724;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).
OS PHOA.
OC Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NRRL 3135 / VAN TIEGHEM / FICUM;
RX MEDLINE=95032134; PubMed=7945393;
RA Ehrlich K.C., Montalbano B.G., Mullaney E.J., Dischinger H.C. Jr.,
RA Ullah A.H.J.;
RT "An acid phosphatase from Aspergillus ficum has homology to
RT Penicillium chrysogenum Phoa."
RL Biochem. Biophys. Res. Commun. 204:63-68(1994).
CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN
CC ALCOHOL + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: TO P.CHRYSOGENUM PHOA AND TO A KLUYVEROMYCES LACTIS
CC ORF (AC P08540).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L20566; AAA62393.1; -.
KW Hydrolase; Glycoprotein; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 417 ACID PHOSPHATASE.
FT ACT_SITE 215 215 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 417 AA; 46735 MW; 896FD2AF541C69C8 CRC64;

Query Match 4.8%; Score 119.5; DB 1; Length 417;
Best Local Similarity 20.6%; Pred. No. 0.2;
Matches 74; Conservative 43; Mismatches 137; Indels 105; Gaps 15;

QY 155 YPKLAPEATRQVNGYIT-----EEVPTLVD---LV 183
Db 26 YPSLAETRAAQATVQPVSPVSNVKGTLNFRVNIWLENTDFDAAAATHELPVLAKKGLLL 85
QY 184 DDEFTTFNSWFCVPGPTNPNRLCALAGTAAGHGKNDLNGIYSSKSIFFAANEKGVSW 243

```

Db 86 NNF-----W--AVTHPEPN-YCRHPLGLDTEGMDNDFFHQIPSNVSTIADLFDTKNIAW 137  
 QY 244 -----LNYDGTNGEFEPDS-----LFFTVNQTGR-SNVVPEVNFQDAYL 283  
 Db 138 GEYQGLPYPGVQGYRYPESGANDYVRNRNPILLFDSVTEALRLRQIKNESFVDDLEN 197  
 QY 284 GVLKFSYINPSCCTNTNSMHPGNSVGEVFKQIYDAIROGPQWDK-TLLITVDET 342  
 Db 198 HRLPYMFITPNM-----TNDGHD-TNITSGDWTWGLFSELLENDYFTKDLIMLTDET 252  
 QY 343 GCGF-----DRVPPPLAVRPDNLTYT-----ETAKNGQRYTLHFDRLGGRMPT 385  
 Db 253 GTYEIGNNIYVTLGGVAPDGLLTKDDTFYTHYSVIASLSTNWGLPSLGRWDCGANLFS 312  
 QY 386 WVI-----SPYKKGYYEIOYCTDPVTKPKAPYSA-TSVLTKL 421  
 Db 313 WLAKKTYGVNVEVDSNLYMNMETHWGLPSLDDDDYSEYAGWPVPTTDASCAGNGLISTV 371

RESULT 10  
 PHOA\_PENCH STANDARD; PRT; 412 AA.  
 AC P37274;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).  
 GN PHOA.  
 OS Penicillium chrysogenum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
 OX NCBI\_TaxID=5076;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=92225342; PubMed=1563629;  
 RA Haas H., Redl B., Friedlin E., Stoeffler G.;  
 RT "Isolation and analysis of the Penicillium chrysogenum phoA gene  
 encoding a secreted phosphate-repressible acid phosphatase.";  
 RL Gene 113.129-133(1992).  
 CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN  
 CC ALCOHOL + ORTHOPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -!- SIMILARITY: TO A NIGER PHOA AND TO A KLUYVEROMYCES LACTIS ORF  
 CC (AC P08540).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M80366; AAA33693.1; -  
 DR FIRM: JN0319; JN0319.  
 KW Hydrolyase; Glycoprotein; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 412 PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE.  
 FT SITE 214 214 PROTON DONOR (BY SIMILARITY).  
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 412 AA; 45628 MW; 5659A7F07BDDFD9 CRC64;

Query Match 4.6%; Score 116; DB 1; Length 412;  
 Best Local Similarity 19.9%; Pred. No. 0.35;  
 Matches 67; Conservative 40; Mismatches 111; Indels 118; Gaps 15;

QY 179 LVLDVDEFTTNSWFSVCPGPTNPNRLCALAGTAAGHKNDLFLNYGISKSIFFEAANE 238  
 Db 75 LSLAKREGILLTNFYA-ISHPSQPN-YCASAG-GDTFGMDNDLFLQIPSNVSTIADLFD 131  
 QY 239 KGYSW-----LNYDGTNGEFEPDSLFTTYV-----NOTSRNVVP--VEN---FF 278  
 Db 132 KHSWEQEDMPAGYQGRYPLSGPNQYRKHNPLVLENSVTDNVRPQIKNFTFY 191  
 QY 279 QDAYLVGLPKFSYINPSCCTNTNSMHPGNSVGEVFKQIYDAIROGPQWDK-TLLFI 337  
 Db 192 DDLKHSLPQHMFITPNM-----TNDADHT-NITVAGNWDVDFLSPLLNKNEYTKDLSVLL 246  
 QY 338 TYDETGGFVDH-----VPPLAVRPDNLTYTETAKNGQRYTLHFDRLGGRMPT 385  
 Db 247 TFDE-GDYSYPNRVFSELVGGALPEHLKGTDDTFY-----HYSIVASISAN 294  
 QY 386 WVI-----SPYKKGYYEIOYCTDPVTKPKAPYSA-TSVLTKL 406  
 Db 295 WGLPSLGRWDCGANLLKMWADKTGYVNWEDTSNVYLNETYPGPMSTDNYSKWPATK 354  
 QY 407 GK-----PAPYSATS 416  
 Db 355 GKCSAGHGTAEVKMYHGLQPTYDASVPVDTVS 390

RESULT 11  
 PHB\_ALCFA STANDARD; PRT; 488 AA.  
 AC P12625;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POLY(3-HYDROXYBUTYRATE) DEPOLYMERASE PRECURSOR (EC 3.-.-.-) (PHB  
 DE DEPOLYMERASE).  
 OS Alcaligenes faecalis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Alcaligenes.  
 OX NCBI\_TaxID=511;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=71;  
 RX MEDLINE=89123016; PubMed=2644188;  
 RA Saito T., Suzuki K., Yamamoto J., Fukui T., Miwa K., Tomita K.,  
 RA Nakanishi S., Odani S., Suzuki J.-I., Ishikawa K.;  
 RT "Cloning, nucleotide sequence, and expression in Escherichia coli of  
 the gene for poly(3-hydroxybutyrate) depolymerase from Alcaligenes  
 faecalis.";  
 RL J. Bacteriol. 171:184-189(1989).  
 CC [2]  
 CC FIBRONECTIN TYPE III DOMAIN.  
 RX MEDLINE=93028390; PubMed=1409594;  
 RA Bork P., Doolittle R.F.;  
 RT "Proposed acquisition of an animal protein domain by bacteria.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8990-8994(1992).  
 CC -!- FUNCTION: THIS PROTEIN DEGRADES WATER-INSOLUBLE AND WATER-SOLUBLE  
 CC PHB TO MONOMERIC D(-)-3-HYDROXYBUTYRATE.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,  
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J04223; AAA21974.1; -  
 DR FIRM: A32235; A32235.  
 DR HSSP: P02751; 1TWC.



Query Match	4.48;	Score 109.5;	DB 1;	Length 729;
Best Local Similarity	21.7%;	Pred. No. 2.2;	Mismatches 145;	Indels 127; Gaps 23;
Matches 88;	Conservative 44;			
QY	110	FNDPBUSVGNL-----EFG-----TTPNNGAIGSGKWADOS	145	
Db	271	FNPADQLISNMTEINLVFSQEIVNVKLTIESATSPPTKIYNNDLASVALVDLN	330	
QY	146	GFNLQALNDYPKLAPREATRQWNGYYTEEVPTVLVDLEFTTFNSWSCVPGGTPNRL	205	
Db	331	NTL-----PYTPAAPRSET---LGFY--PWLPF-----KPTQRYLLSCTRNLNLP----	371	
QY	206	CALAGTAAGHKNDODFLNYGISGSKSFIAANEKGVSMLNYDG---TNG--EFEPDSLFF	260	
Db	372	-----TYTGQSQQITDSIQTLGHSDIMFYFIENAVPIHLRTDGEFTGIYHFOTKPLK	426	
QY	261	TYNNTSRSNVVPVENFFQ-----DAYGLVKPFYSINPSCGTNT-NSMHPTGNVSYGE	314	
Db	427	THSWOTNLSGLGPKKLLTPTTEGDHPHGPLP-----AANTRGYHOTLNINSYTE	476	

QY 315 VFVKQIVDAIROGP-OWDKTLLFTYDEGGFYDHPVPLAVRPDNLTYTETAKNGO-KY 372  
 DB 477 A-----TAIRPAQVGYNTYPMNFESYNGGPELPVPTA-----DTQYNDDPENGAIRF 525  
 QY 373 TLHPDRIGRMP-----WVISPYSKGYI--EQYG-----TPDVT 406  
 DB 526 TMGYOH--GOLTTSSOELERYTFNPOSKGRAPKQOFNOQAPLNLENTNGTLLPSPDIP 583  
 QY 407 GKP-----APYSA---TSVLKTLGYLEWIE---DFTPRV 434  
 DB 584 GKPMHFMNTLNTYGPLTALNNTAPFPNGQIMDKELDTDLKPL 628

## RESULT 15

CCAA\_BACTU STANDARD; PRT; 1257 AA.  
 AC Q45754;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PESTICIDIAL CRYSTAL PROTEIN CRY12AA (INSECTICIDAL DELTA-ENDOTOXIN  
 DE CRYXIIA(A)) (CRYSTALLINE ENTOMOCIDAL PROTOXIN) (142 KDA CRYSTAL  
 DE PROTEIN).  
 GN CRY12AA OR CRYXIIA(A) OR CRYVB.  
 OS Bacillus thuringiensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1428;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.  
 RC STRAIN-NRRL B-18244 / PS33f2;  
 RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,  
 RA Sick A.J.;  
 RT "Novel Bacillus thuringiensis microbes active against nematodes, and  
 RT genes encoding novel nematode-active toxins cloned from Bacillus  
 RT thuringi."  
 RL Patent number EP0462721, 27-DEC-1991.  
 CC -!- FUNCTION: ENDOTOXIN WITH NEMATICIDAL ACTIVITY.  
 CC -!- DEVELOPMENTAL STAGE: THE CRYSTALL PROTEIN IS PRODUCED DURING  
 CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
 CC OF THE SPORE COAT.  
 CC -!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
 CC N-TERMINUS.  
 CC -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 CC EMBL; L07027; AAA22355.1; ..  
 DR InterPro; IPR001178; Endotoxin.  
 DR Pfam; PF00555; endotoxin; 2.  
 KW Toxin; Sporulation.  
 SQ SEQUENCE 1257 AA; 142265 MW; 3D988FFC6C0E3981 CRC64;

Query Match 4.3%; Score 107; DB 1; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 6.8; Indels 166; Gaps 23;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
 QY 6 LITGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESINLAKDKVENIVWLLEN--- 61  
 DB 74 VLSVLTLPVGVTAASASTIVSIFWPKIFGDKPNAKNIFEELKPOIALIQODITNYQD 133  
 QY 62 ----RSDNLLGVRRQGLDNPINNPGFCNKNASDPSSGKYCTQAKDYDSVFNQ---- 112  
 DB 134 AINOKKFDLSQKTLNLTV--AIDNNDYVTAK-----TOLENLNLSTLTSISIF 180

QY 113 -PDHSVTGNLLEFYCTYTPNNGAIAASKYV-ADQSGFLNAQLNDYKPLAPEATROVNGY 170  
 DB 181 IPEGYETG-GLPYAMVANAHILLRLDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234  
 QY 171 YTEEEVPTLVLDLVEFTTF--NSWFSCVPGTPNPNRLCALAGTAAGHGKNDODDFLNYGIS 228  
 DB 235 HTEAVIKAFENGLDKFKSLDVNST-----NKKANY--- 264  
 QY 229 SKSIFEAANEKGVSWLNYDGTNGEFPDLSLFTYVN-----QTSRNVVVPVENF 277  
 DB 265 IKGMTKMLDLVALMPTDPDHYOKEVEIEFTRTISSPIQVPVKMONTSSSIIVPSDLF 324  
 QY 278 FQDAYLGLPK--ES-----YINPSCGT-----NTNSMHPGTN 309  
 DB 325 H---YQGLVKLEFSTRDNDGLAKFTGIRNTFFKSPNTHYHYHDFSYNTQS---SGN 378  
 QY 310 VSYG-----EVFVKQIYD-----AIRGQPDWKTLFF 336  
 DB 379 ISRGSSNPIDLNNPIISTCIRNSFYKALAGSSVLVNFKDGTOGYAFAQAP----- 430  
 QY 337 ITYDEGGFYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLHEDRLGGRMPTWVI 388  
 DB 431 -----TGGAWDSFIESDCAPEGHKLNYYTSPGDTLRDFINVTLLISTPTINELSTEKI 485  
 QY 389 SPY-SKKGYYEQYGTDPVTGKP 409  
 DB 486 KGFPKGYIKNOGIMKYGKP 507

Search completed: December 4, 2001, 15:08:21  
 Job time: 81 sec







Qy	224	NYGISSKSIPEAANEKGVSWLNDGTNGEFGEDPSLFTFYVYNQTSRSNVVPVENFFQ----	2719
Db	182	VRGFQPTVFTELESSEGTGFE-----GIYQSPFNCLEFYRNMRLKY-----VDNPFQYHLS	232
Qy	280	---DAYLGVLPKFSYINP---SCCGTNTNSMHTPGNSYGEVFKQIYDAIRQGPWDKTT	333
Db	233	KFRCKBGGKLPNVVIEPRYFKILSAPANDDHPKNDVVVEGONLVKEIYEALRASPOWNEI	292
Qy	334	LLFTYDDEGGFYDHHVPPPL--AVRPDLNTYTETAKNGQKXTHLHEDRLGGRMPTWVLSPY	391
Db	293	LFVYDEHGGYIDHVPYVIGVNPNDGLVGPE-----PINFKFDRLGVRVPALLISPW	346
Qy	392	SKKGYI--EYQGTDPVTGKPAYSATSVLTKLGLYLDIEDF--TPRVAHSPSFDHLIG--TT	447
Db	347	IEPTGVLIHPENGPEPT----SQFESSHSATPLKIKFNLSFLTKRDMAGTLDVAINRTS	402
Qy	448	LREDAPAL 456	
Db	403	PTDCPVTLL 411	
RESULT	2		
Q9SRQ7			
ID	Q9SRQ7	PRELIMINARY;	PRT; 538 AA.
AC	Q9SRQ7		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)		
DE	01-MAY-2000 (TEMBLrel. 13, Last annotation update)		
DE	T21P5.5 PROTEIN.		
GN	T21P5.5.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.		
NCBI	taxid:3702;		
ON	NCBI_taxid=3702;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,		
RA	Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,		
RA	Bowman C.L., White H., Nierman W.C., Fraser C.M.;		
RT	"Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence."		
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
RF	EMBL; AC009895; AAF01582.1; -		
SR	SEQUENCE 538 AA; 60723 MW; B559583CB39ACE6 CRC64;		

Query Match	18.3%;	Score 458;	DB 10;	Length 538;
Best Local Similarity	30.1%;	Pred. No. 3.7e-27;		
Matches 131;	Conservative 73;	Mismatches 171;	Indels 60;	Gaps 16;
QY	51	VENTVWLILENSFDNLTGVR--RQGLNDPNIINGPFQNYKNASDPSSSKYCTQAKYD	108	
Db	14	IKTVILVQENSFDHLGFWFKELAREIDGVTKSPKSNWSSSDINSLR--VVFQDSQ	71	
QY	109	VFN-DPDHVSVTGNLIFY-----GTYTPNNGAIAISGKVADQSGFL-NAQLNDYPKLAP	160	
Db	72	VYVPDPGHSIQDIEQVFGKMPDSCGKPDNPQH-----PNMSGFAQAERN-----K	118	
QY	161	EAATQVMGYIIEEBEVTVLDVDFETTFNSWSCVPGTNPRLCALACTAAGHKND	220	
Db	119	KGMSAVAMGFKPALPVYKELQVAFCDRWFAFSPASTQPNRLVHSAATSHGATSNK	178	
QY	221	DFLNYGISKSIFEAANEKGVSMWLDGTNGEPEPDSLFF-----TYYNQTSRNVVPV	274	
Db	179	KLLLEGPPQKTFPESLDEAGFSR---GIYYQPPSTLFLYRLNKLKYLTHFQYGI---	231	
QY	275	ENFPDQAYLGVLPAFESYNP---SCGTNTNNSMHPGTNVSGVGVFKQIYDAIRGQPWD	331	
Db	232	-OFKKCKGKGLPNVVMVEQRMFDLLSTPANDQHPSHDVSGOKLYKEVYEALRSSPWN	290	
QY	332	KTLFLFYDETGGFDVHPPLAVRFDNLVLTETAKNGOKYTLFLDRLGRMFTWTWSPY	391	

Db	291	ELIFIIYVDHGFGYDHWPTPV-----DGVNPDGILGPPPNFEFNLGRVVRVTFEISFW	346
Qy	392	SKKGVIQYGTDPVTKGKAP-----YSATSVLKTLGILWLDIEDF-TPRVAHSPSFDHLI-	444
Db	347	IBPGTV-----IHGNGPYPRSQYEHHSIPATVTKTFIKLKDPLSKROSAGTFESVIT	399
Qy	445	GTTREDAPALAKTP 459	
Db	400	RSPQDCPETLSTP 414	
RESULT	3		
Q9LML5	ID	Q9LML5 PRELIMINARY; PRT; 533 AA.	
AC	09LML5;		
DC	01-OCT-2000 (TEMBLrel. 15, Created)		
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TEMBLrel. 15, Last annotation update)		
DE	F10K1.5 PROTEIN.		
GS	F10K1.5.		
OS	Arabidopsis thaliana (Mouse-ear cross).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Liu S.X., Chan A., Yu G., Etgu P., Lee J.M., Lenz C., Pham P.,		
RA	Sakano H., Toriumi M., Chung M., Goldsmith A., Liu A., Smith A.,		
RA	Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,		
RA	Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B.,		
RA	Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,		
RA	Ecker J.R., Federspiel N.A., Theologis A.;		
RT	"The sequence of RAC F10K1 from Arabidopsis thaliana chromosome 1.";		
RN	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Theologis A.;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Theologis A.;		
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Theologis;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AC067971; AAF82197.1; "		
SO	SEQUENCE 533 AA; 60038 MW; 607ADA226E2E5215 CRC64;		
Query Match	18.1%;	Score 453.5; DB 10; Length 533;	
Best Local Similarity	30.9%;	Pred. No. 8e-27;	
Matches 139; Conservative	77; Mismatches 167; Indels 67; Gaps		
Qy	34	FGYKGGKSEIE-----NLKKVENIWLILENSFDNILGVVR--RQGLDNPINN 83	
Db	13	FCYLLISSQSTEFKNSQPKHKTQGIPTKIVVVVNNFSDHILGWLKSRTPEIDG--ITG 70	
Qy	84	PCNTKNASDPSSGCKYQAKDYDSVF--NDPDHSTVGNLLEFYGTYPNNGAIAAGKV 141	
Db	71	KESNPLAVSDPNSKIFVSD--DAVFVMDPGHSFOAIRQEIFG----SNDTSGDKM- 122	
Qy	142	ADQSGLNAQLNDYPKLAPEATQVMGYTTEVEYPTLVLDVDEFTTFNSWFSVCGPTN 201	
Db	123	---NGFAQQSSEMEPGMA-----KNWSGPKPEVLPTVYTELANEFGVDFRWFASTTQ 174	
Qy	202	PNRLCALGTAAGHGKNDDDLNTYGISKSIFFEAANEKGVSWLNYDGTNGFEFPOSLEFT 261	

Db 175 PNRVYHSATSHGCSNVKDLVGFPOKTIFDSDLNGLSE-----GIYYONIPATFFFK 230  
 QY 262 YNQTNRSNVYPVEN-----FPQDAYLGVLPKFSYINPSCCGTN---TNSMHPGTVNSYGE 314  
 Db 231 SURRL--KHLVYFHSYALKFKLDAKGLKLPNTSVVEQRYFDLDPANDRHPSHDVAAGQ 288  
 QY 315 VFVKQIYDAIRQGPQMDKTLITFYDETGGFYDHVPPPL--AVRPDNLITYTETAKNGQY 372  
 Db 289 RFVKEVETLRSSPOMKEMALLITYDEHGGFYDHVPTPVKGVNPDGII-----GPDPE 342  
 QY 373 TLHFDLGRMPTWVIVSYKKGYY--EQYCTDPTVCKPAPYSA---TSVLKTLGLYLDI 427  
 Db 343 YFGFDLGRVVRPTFLISWIEKGTVIHEPEG-----PPHSQFEHSSIPATVKLFNL 395  
 QY 428 ED--FTPRVHSPSFDHL--IGTTLREDAP 453  
 Db 396 KSHFLTKDAWAGTFEYFRIIRDSRQDCP 425

RESULT 4  
 Q9S816 PRELIMINARY; PRT: 521 AA.  
 AC Q9S816;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DE PUTATIVE PHOSPHOLIPASE.  
 GN T12J13.18 OR T21P5.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
 RA Roming C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence."  
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC003327; AAF03477.1; --  
 DR EMBL; AC003895; AAF01581.1; --  
 SQ SEQUENCE 521 AA: 59071 MW: 1A39ACEP6807B8EF CRC64;

Query Match 17.58; Score 437; DB 10; Length 521;  
 Best Local Similarity 30.08; Pred. No. 1.5e-25;  
 Matches 133; Conservative 73; Mismatches 161; Indels 76; Gaps 18;

QY 51 VENTVWLILNRSFONILG-----GVRRO-----GLDNPIN--NGPFCNYKNAS 92  
 Db 13 IKTIVLVQENSFDPHTLIGWPKELNREIDGVKMSQKPNPFGSDLNHNHVVFGDSQYV 72  
 QY 93 DPSSGKVCYQAND--YDSVNDPDHVSYTGNNLEFYCTYTPNNGAATASGVVADQSGFL-NA 150  
 Db 73 DPNPGH---SIRDIETQYEGKPMDS-----GHDPNPGP-----ATMSGFAQNA 113  
 QY 151 QUNDYPKLAPEATRVQVMGYTTEEVPTLVLDVDEFTTNSWFSVCGPPTNPNRICALAG 210  
 Db 114 ERK-----MKGSSAVMNGFKPDALPVYKELQVFAICDRWFASVPGATQPNRLFIHSA 167  
 QY 211 TAAGHKNDLFLNYGKSSIFEAANEKGVSWLNYDGTNGCEPEPSLFTYVNO---TS 267  
 Db 168 TSHGTTNERNKLLIEGFPQKTFIPESLDEAGTFE---GIYQCPPTLYRNRLKLYLT 223  
 QY 268 RSNVVPVENFPQDAYLGVLPKFSYI-----NPSCCGNTNNSMHPGTVNSYGEVFK 318  
 Db 224 RFHYDGLQ-FKDKCEGNLPNVVVEQRYWYDLLNPA-----NDDHPSHDVSEGOGLVK 276  
 QY 319 QIYDAIRQGPQMDKTLITFYDETGGFYDHVPPPLAVRPDNLITYTETAKNGQYTLHFDH 378  
 Db 277 EYVIALRSSPQWNEILLITYDEHGGFYDHVPTPL-----DGVNPDGILGPPYPNFEFR 332

QY 379 LGRMPTWVIVSYKKGYYEQYCTDPTVCKPAPYSATSVLKTGLYLDIEDF--TPRVAHS 437  
 Db 333 LGVRVPTTFISPIWIEPCTVHLGNSGPLY--MSQYEHSSIPATVKKIPKDKFLTKRDSWA 390  
 QY 438 PPSDHLI--GTTLRREDAPIALKTP 459  
 Db 391 GTFESVITRNSPRODCCPETLSNP 413

RESULT 5  
 Q9SMN8 PRELIMINARY; PRT: 462 AA.  
 AC Q9SMN8;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DE HYPOTHETICAL 51.3 KDA PROTEIN.  
 GN T8P19.120.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choise N., Robert C., Brottier P., Wincker P., Cattolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project.  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL133315; CAB62350.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 462 AA: 51284 MW: EE6492B529991095 CRC64;

Query Match 17.44; Score 436; DB 10; Length 462;  
 Best Local Similarity 32.08; Pred. No. 1.5e-25;  
 Matches 125; Conservative 61; Mismatches 165; Indels 40; Gaps 13;

QY 78 NPINN---GFCNYKNASDPSSGKCYQAKYDVSFNDPDHVSYTGNNLEFYCTYTPNNGA 134  
 Db 6 NPTINGVTGECN----PVPNSTQTICFTSDAEFVDPDPGHSPFEAVEQVFGS--CGFQ 58  
 QY 135 IASGVVADQSGFLNAQLNDYKLAPEEATROVMGYTTEEVPTLVLDVDEFTTNSWFS 194  
 Db 59 IPS-----MKGFEVQALS-----MFCNLSEYVMKGFPEAVPYAELVKEAFVDRWS 107  
 QY 195 CVPGPTNPNRICALAGTAAGHKNDLFLNYGKSSIFEAANEKGVSWLNYDGTNGEPE 254  
 Db 108 SIPGPTQPNRLFVYSATSHGTSYKVKLAQGYPOKTFPSLSHNSNIDF----GIYFONI 163  
 QY 255 PDSLEFTYVNOTSR--SNVVPVE--NFODAYLGVLPKFSYINP---SCCGTNTNSMHPG 309  
 Db 164 PTTLYRNRLQKLYIFNLHQLKFKDAKGLPFLVIEPRYFDLKLGLPANDHPSHD 223  
 QY 310 VSYGEVFPVQITDARTQCPQMDKTLITFYDETGGFYDHVPPPLAVRPDNLITYTETAKNG 369  
 Db 224 VANGQKLVKEVYEARLSSPQWNETLLVITYDEHGGFYDHVKTYPYVGINPDGNTGPAGCF 283  
 QY 370 QKTYLHFDLGRMPTWVIVSYKKGYYEQYCTDPTVCKPAPYSATSVLKTGLYLDIED 429  
 Db 284 FK----FDRLGVRVPTINVSPIQKGVVSEAKGPT--ESSEYEHSSIPATIKKLFNLSS 337  
 QY 430 --FTPRVHSPSFDHLIG--TTLREDAPIAL 456  
 Db 338 NFLTHRDAAWATPEDVSVSHLTTPRTDCPMTL 368

RESULT 6



RY MEDLINE-99454884; PubMed-10523590;  
RX Korbsrisate S, Swanasai N, Leelaporn A., Ezaki T., Kawamura Y.,  
RA Sarasombath S.;  
RT "Cloning and characterization of a nonhemolytic phospholipase C gene  
RT from *Burkholderia pseudomallei*,"  
RL J. Clin. Microbiol. 37:3742-3745(1999).  
RL EMBL: AF107252; AAF17299.1; --  
SQ SEQUENCE 700 AA: 771910; -- 5E2AE6FCDF0BF129B CRC64;

Db 68 GTL--SGTDGNTVSLFQKGWNP-----MTQALDATGYTMPYRFDTRGRPFLDGACV 119

Qy 111 NDPDHSVTGNLEFFYCTGYTPNNGAIASGKVVDQSGFLNAQINDYBPKLAPPEATRQWNGY 170

Db	120	NDPHS	-----WYAMHESNGGV-----NDNWLPDQ-----AKTISAHPTVMGY	160
QY	171	YTEEEVTLVDLVDEFTTNSWFCSPGPTNPNRLCALAGTAGHGKNDDEL	223	
Db	161	YTRQDIPHYLLADAFTVCDYFCSVLGGTLPNRLYLSATIDPDQNGGPELOSP	220	
QY	224	---NYGISSSKSIFFAAANEKGVSWLYNDCTNGCFEPDLSLFTVYNQTS---	267	
Db	221	VRRFGWRIMPQLNSDA---GVSMKYV---RNTLGLPTSSVLVTGSLVTSFKQSADPRSDL	274	
QY	268	-RSNVVP--VENFFODAYLGLVLPFSYINPSCCGTNTNSMHPGTNYSYGEVFVKQIYDAI	324	
Db	275	VRFGVAPSPASFAADVLNLRPLRVSMVTPNV---LESEHPAVPAAGAFAINILRL	330	
QY	325	RGQPO-NDKTLFLFYDEYTGCFYDHYPPPLVAYRPDNLTYTETAKNGKYL-HFDR---	378	
Db	331	LANPAVWEKTLALVSDNGGFFDHHVPATA-----PAGTPGEYVTPDIDQVPGS	381	
QY	379	-----LGGRMPTWVISPSYK	393	
Db	382	GGIRGPGLGRFVPCFVISPYSR	404	

OS *Mycobacterium tuberculosis*.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;

11	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RC	MEDLINE=9825987; PubMed=9634230;
RX	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.
RX	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.
RA	Gordon S.V., Eiglmeter K., Gas S., Barry C.E. III, Tekaiia F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moute S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.
RA	Taylor K., Whitehead S., Barrell B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence".
RT	Nature 393:537-544 (1998).
DR	EMBL; Z83860; CAB06147.1; -.
DR	Tuberculist; RV2350C; -.

```
KW Complete proteome.
SQ SEQUENCE 512 AA; 55623 MW; B82672FA0C81DA9B CRC64;

Query Match 11.5%; Score 288; DB 2; Length 512;
Best Local Similarity 26.1%; Pred. No. 4.6e-14;
Matches 141; Conservative 60; Mismatches 184; Indels 156; Gaps 29;

QY 14 ASLGASPVTSYTSVRAPFGYKPGSKESIEHLKDKVENIVWLLIENSDNLGGVR- 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 AFMSLAGFIIEKAYGA-----GPCPG-----HLTD-IEHLVLLMQENRSDHFGTSLD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 73 RQGLD-----NPTNNGPFCNYKNASDPSSGKYCTQAKDYDSV-----F 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TRGDDTTTPPVFAQSGWNPT-----QAVDPAG-----VTLPRYFDTTRCPVAGCV 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 111 NDPHSVTGNLLEFYGTTPNNGAIAAGKVVADQSGFLNLAQNDYKFLAPEATRVQWGY 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 NDPDHSWIGHNSW-----NGGA-----NDNWLPAQVPFSPLOGNPEVT---MGF 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 171 VTEEVPTLVDLVEFTTFNSWFSVCPGPTNPNRLCALA-----GTAAGH---GKNDDD 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 YTRDLPIHLLADFTVCDGVCFLSLGGTTNRLYKMSAWIDPDGTDGSPVLIBPNIQP 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 222 FLNIGISKSIFPANEKGVSW-----LNYDGTNGEF-----EPDSLFTY 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 LQHY--SWRTIMPENLEDAGSVKQYQNKLLGALNNTVVGYNGLVDFKQADPRS----- 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 263 VNQTSRSNVVPVE--NFTQDAYLGVLKFSYINPSCCGTNTNSMHPGTNWSYGEVFKQI 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 --NLARGISPTPLDFAADYRNRLPKYSWVLPGF-----LLSEHPAFVNVGAV---AI 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 321 YDAR---QCPQ-WDKTLFTYDETCGYDHYVPP-----LAVRPDLNLTYTETAK 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 VDALKRILLSNPVWKEKTALIVYDENGGFHDVPPPTPPGTFGEVTPD--IDSVPGS 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 368 NGQKYTLHFLRGRMPTWISPYSKKGYIEQYGDTPVT-----GKPAPYSATSV 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 GGIRGPI---GLGFRVPCLVISPY-SRGPLVHDTFTSTLKLIRARFGVPVPLTAWR 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 418 LKTLGLYLDIEDR-IPRVAHSPSDH-----LGLTLREDAPIALKTPH 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 DATVDGTMSTFNFAAPPNPSKPNLDHPRNLALPKLPQCVPNVNLGVTYKTAIPYRVPFP 494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 461 T 461
   :
Db 495 S 495

RESULT 11
ID Q50560 PRELIMINARY; PRT; 520 AA.
AC Q50560; Q50771; O08223;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (T-EMBLrel. 14, Last annotation update)
DE PHOSPHOLIPASE C.
GN PLCA OR RV2351C OR MTCY98.20C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Johansen K.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 9-520 FROM N.A.
RC STRAIN=H37RV;
RA del Portillo P., Murillo L.A., Patarroyo M.E.;
RX MEDLINE=91372944; PubMed=1909099;
RA Parra C.A., Londono L.P., del Portillo P., Patarroyo M.E.;
RT "Isolation, characterization, and molecular cloning of a specific
```

```
Myobacterium tuberculosis antigen gene: identification of a species-
specific sequence.";
RL Infect. Immun. 59:3411-3417(1991).
RN [3]
RP SEQUENCE OF 9-520 FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=92042637; PubMed=1939567;
RA del Portillo P., Murillo L.A., Patarroyo M.E.;
RT "Amplification of a species-specific DNA fragment of Mycobacterium
tuberculosis and its possible use in diagnosis.";
RL J. Clin. Microbiol. 29:2163-2168(1991).
RN [4]
RP SEQUENCE OF 9-520 FROM N.A.
RC STRAIN=H37RV;
RA Leao S.C., Rocha C.L., Parra C.A., Patarroyo M.E.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; U49511; AAC18943.1;
DR EMBL; L11868; AAB59164.1;
DR EMBL; Z83860; CAB06148.1;
DR TubercuList; RV2351c;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 520 AA; 56138 MW; 5844EF0C064288A7 CRC64;

Query Match 11.4%; Score 286; DB 2; Length 520;
Best Local Similarity 28.6%; Pred. No. 6.7e-14;
Matches 128; Conservative 40; Mismatches 161; Indels 118; Gaps 20;

QY 7 LTGGLGLASL-GLASPVTSYTSVRAPFGYKPGSKESIEHLKDKVENIVWLLIENRSD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 LTGAGAAFLMDWAAPVIEKAYGA-----GPCPG-----HLTD-IEHLVLLMQENRSD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 66 NILGGVRRGDLNPNNGPFCNYKNASDPSSGKY-----CTQAKDYDSV----- 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 HYFGLTSS-----TNGFNAASPAFQMGWNPMTQALDPAGVTIPERLDTTRGP 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 -----FNDDHSVTGNLLEFYGTTPNNGAIAAGKVVADQSGFLNLAQNDYKFLAPEA 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 FLDGECVNDPEHQQWGMHLAW-----NGGA-----ND--NWLPAQA 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 164 TRQ-----VMGYTTEEVPTLVDLVEFTTFNSWFSVCPGPTNPNRLCAL-----AG 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 TTRAGPVVPLTMGYTYTRDIPHYLLADFTTICDGYHCSLLGTGLPRLWLSANIDPAG 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 211 TAAGHKNDLDFLNGISKSIF-EAANEKGVSWLYNDGTN-GEFEPDLSLFTTYVQTSR 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 TDGGPQLVPEFGFLPQQFSWRIMPENLEDAGSVKQYQNKLLGALNNTVVGYNGLVDFKQAFR 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 269 SNVVPVEN-----FFQDAYLGVLKFSYINPSCCGTNTNSMHPGTNWSYGEV 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 QAADPRSNLARXYGIAPTYPGDFDAADVRLNPKVSWLVPNI---LQSEHPALPVALGAV 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 316 FVKQIYDAIRQGPQ-WDKTLFTYDETCGYDHYVPPPLA-----VRPNLIVTETA 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 SWVTALRILLSNPVWKEKTALIVYDENGGFHDVPPPTPPGTFGEVTPVFNID-AVPG 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 367 KNGQKYTLHFLRGRMPTWISPYSK 393
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 383 SGGIRGLP---GLGPRVPCIVISPSYR 406

RESULT 12

Q50561 PRELIMINARY; PRT: 521 AA.

AC Q50561

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-NOV-1996 (Tremblrel. 01, Last annotation update)

DE PHOSPHOLIPASE C.

GN MPCB.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RA Johansen K.A.;

RC STRAIN-H37RV;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U49511; AAC18944.1; -

SQ SEQUENCE 521 AA; 56424 MW; 7C915196E6914256 CRC64;

Query Match 10.8%; Score 271; DB 2; Length 521;

Best Local Similarity 25.5%; Pred. No. 9.7e-13;

Matches 138; Conservative 60; Mismatches 186; Indels 158; Gaps 29;

QY 14 ASGLASPVSEYTSVREAPFGYKPGSKESIEENLKDKVENIWLLENSEDNILGGVR- 72

Db 27 AFMSLAGPIIERAYCA-----GPCPG-----HLTD-IEHIVLLMOENRSPDHFGLTSD 74

QY 73 RQGLD-----NPTNNGPCFNKASDPSSGKCYTQAKYDSV-----F 110

Db 75 TRGDFTTPPVFAOSGWNPMT-----QAVDPAG---VTLPRYFDTTRGPLVAGECV 123

QY 111 NDPHSVTGNLEFYGTYPNNGAIAAGKVVADQSGFLNAOLNDYKPLAPEATQVMGY 170

Db 124 NDPHSWIGMHSW-----NGGA-----NDNWLPAQVPFSPLOGNVPVT---MGF 165

QY 171 YTEEVPTLVLDVDEFTTNSWFSVCPGPTNPNRLCALA-----GTAAGH---GKND 221

Db 166 YTRDLPFIYLLADFTVCDYFCSLGGTTPNRLYWNAMIDPDGTCGPVLIENPNIQ 225

QY 222 FLNGYSSKSIPEANEGYSW-----LNYDGTNGEF---EPDSLFTFY 262

Db 226 LOHY--SWRIMPENLEDAGVSKYONKLLGALNTVVGYNGLVDFKQADPRS----- 278

QY 263 VNQTSRNVVPE--NFFQDAYLGLVPKFSYINPSCCGTNTNSMHTGNGVSYGEVFKQI 320

Db 279 --NLARGISPTYPLODFAADVNRNLTYSWVLPGF---LLSEHPAFVNVGAV---AI 329

QY 321 YDAIR---QGPQ-WDKTLFFTYDEFGFYDHPVP-----LAVRPDLNLTETAK 367

Db 330 YDAKILLNSPAPVNEKTAALVNDENGFFDHPVPPPTPGTGPGEVTVPD---IDSVPGS 387

QY 368 NGQKTYLHFDRLGGRMPTWVISPYSKKGIEQYGTDPVT-----GKPAYSATSV 417

Db 388 GGIRGAI---GLGPRVPCLEISPYT--GPLWHDFTDHTSLKILRAREGVVPVNLTAWR 442

QY 418 LKTYLWLDIEF--TPRVHSPSPDH-----LIGTTLREDAPALKTPT 459

Db 443 DATYGDITSTFNAAPPNPSKPNLDHPRAQCALPKLPQCVPNVAVLGTVTTKTAIPRVFPF 502

QY 460 HT 461

Db 503 QS 504

RESULT 13

Q9A412 PRELIMINARY; PRT: 686 AA.

ID Q9A412

AC Q9A412;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE PHOSPHOLIPASE C.

GN CC3031.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI\_TaxID=69394;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C.; Feldblyum T.V.; Laub M.T.; Paulsen I.T.; Nelson K.E.;

RA Eisen J.; Heidelberg J.F.; Alley M.R.K.; Ohta N.; Maddock J.R.;

RA Potocka I.; Nelson W.C.; Newton A.; Stephens C.; Phadke N.D.; Ely B.;

RA DeRoy R.T.; Dodson R.J.; Durkin A.S.; Gwinn M.L.; Haft D.H.;

RA Kolonay J.F.; Smit J.; Craven M.B.; Khouri H.; Shetty J.; Berry K.;

RA Uterback T.; Tran K.; Wolf A.; Vamathevan J.; Ermolaeva M.; White O.;

RA Salzberg S.L.; Venter J.C.; Shapiro L.; Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL: AF005965; AAK24993.1; -

DR TIGR; CC3031; -

KW Complete proteome.

SQ SEQUENCE 686 AA; 75591 MW; 00EF60FAA00F8AFF CRC64;

Query Match 10.5%; Score 263.5; DB 2; Length 686;

Best Local Similarity 24.8%; Pred. No. 5.5e-12;

Matches 129; Conservative 77; Mismatches 186; Indels 129; Gaps 26;

QY 5 ALTLTGGLIA---SIGLASPVSEYTSVREAPFGYKPGSKESIEENLKDKVENIWLLEN 61

Db 8 SLLAALGALPALPAAAAIDADVRS-----GTIQD-VEHVILMOEN 50

QY 62 RSPDNILG---GVRQGLDNPINNGPCFNKASDPSSGKCYTQAKYDSVFNQD-PDHSV 117

Db 51 RSPDHFYGLNGVRFGDRPFI-----PVRDAAGKRESSVVFQVQMSKOKLLAPPLNSA 104

QY 118 -TGNNLEFYGTYPNNGAIAAGKVVADQSGFLNAOLNDYKPLAPEATQVMGYTEEV 176

Db 105 QTFAHMRVEG---TPHSMTDA-----QDAWQDQGRMDR---PDAKKPMSMGYFQADI 151

QY 177 PTLVDLVDEFTTNSWFSVCPGPTNPNRLCALAGTAAGHGK-----NDDDFLNYGIS 228

Db 152 PFQALADAFITLCAHCHSTGTCTNTNRLFLWTGNDLCKAGGSPISNSHDFAEKGA 211

QY 229 SKSIF-----EAAKNGKVSUNYDGTNGEPEPDSL--FFTY-----VNQTSR 268

Db 212 KESYTWTTTPPERLLQAGYSWRIYQDMADNFTONPLAGFKAYRDAYKDLFGSPNRLKQLG 271

QY 269 SNVYPVNFQDAYLGLVPKFSYINPSCCGTNTNSMHTGNGVSYGEVFKQIYDAIRQG 327

Db 272 S-TWHLDKLREDVNGRLPQVSWI-----TAPADSEHPNPSPPAQGADYATRIDALTAD 326

QY 328 PQ-WDKTLFFTYDEFGFYDHPVPPPLAVRPD-----NLTYYTET 365

Db 327 PKYWARVFLVWFVNDENGFFDHPVPPAPPSPSYDASGLKLLGASTVDLTAEHHRVNRNTEARS 386

QY 366 AKN---GQKTYLHFDRLGGRMPTWVISPYSKKGIEQYGTDPVTGKPPAPYSATSVLKTGL 422

Db 387 ERDLMGRPY-----GLGPRVPLYVISPMSRGVWNS-----QVFDHTSVLRELE 431

QY 423 YLWDIEDFTPRVAHSPSPDHLIG--TTLREDAPIALKTPT 461

Db 432 QRFGVME--PNI--SPWRRAVCGDLTTCFD-----FKTPNT 463

RESULT 14

Q9L5C5 PRELIMINARY; PRT: 577 AA.

ID Q9L5C5

AC Q9L5C5;

Query Match 8.1%; Score 203.5; DB 2; Length 514;  
Best Local Similarity 21.7%; Pred. No. 1.6e-07;  
Matches 117; Conservative 68; Mismatches 156; Indels 199; Gaps 27;

QY 10 LGLLASGLASPTSEYTSVREAPFGYKPGSKESIENLKDK-----VENIVW 56

[illegible]

RESULT	15
ID	Q47936
PRELIMINARY;	PRT; 514 AA.
AC	Q47936;
DT	01-NOV-1996 (TREMBlrel. 01, Created)
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT	01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE	ACID PHOSPHATASE PRECURSOR.
OS	Francisella tularensis.
OC	Bacteria; Proteobacteria; gamma subdivision; Francisella group; OC Francisella.







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2001, 15:07:00 ; Search time 15.89 Seconds  
(without alignments)  
657.113 Million cell updates/sec

Title: US-09-426-072-2  
Perfect score: 2501  
Sequence: 1 MKSTALLTGLGLASGLAS.....GTTLRDAPIALKTPHTFSV 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2.6/ptodata/2/iaa/PCUS\_COMB.pep.\*  
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2501	100.0	464	US-09-426-072-2	Sequence 2, Appl
2	111.5	4.5	1041	US-08-494-714-2	Sequence 2, Appl
3	111.5	4.5	1041	PCT-US96-10782-2	Sequence 2, Appl
4	109.5	4.4	579	5223424-13	Patent No. 5223424
5	107	4.3	1257	US-08-049-783-2	Sequence 2, Appl
6	107	4.3	1257	US-08-158-232-6	Sequence 6, Appl
7	107	4.3	1257	US-08-304-626-6	Sequence 6, Appl
8	107	4.3	1257	US-08-316-301A-6	Sequence 6, Appl
9	107	4.3	1257	US-08-611-928-6	Sequence 6, Appl
10	107	4.3	1257	US-09-173-891-6	Sequence 6, Appl
11	107	4.3	1257	US-09-076-137-6	Sequence 6, Appl
12	107	4.3	1257	PCT-US92-03624-6	Sequence 6, Appl
13	104.5	4.2	737	US-08-119-361-5	Sequence 5, Appl
14	104.5	4.2	737	US-08-336-308A-4	Sequence 4, Appl
15	104.5	4.2	737	US-08-822-324-4	Sequence 4, Appl
16	104.5	4.2	737	US-09-490-931-4	Sequence 4, Appl
17	104.5	4.2	1687	US-08-570-311-29	Sequence 29, Appl
18	104.5	4.2	1704	US-08-336-308A-10	Sequence 10, Appl
19	104.5	4.2	1704	US-08-822-324-6	Sequence 6, Appl
20	104.5	4.2	1704	US-09-490-931-10	Sequence 10, Appl
21	103.5	4.1	1290	US-08-470-350B-2	Sequence 2, Appl
22	102	4.1	433	US-09-104-623A-4	Sequence 4, Appl
23	102	4.1	635	US-08-873-479-43	Sequence 43, Appl
24	101.5	4.1	746	PCT-US95-10509-2	Sequence 2, Appl
25	97.5	3.9	713	US-09-346-237-9	Sequence 9, Appl
26	96.5	3.9	481	US-08-943-714-10	Sequence 10, Appl
27	96	3.8	532	US-08-579-777A-2	Sequence 2, Appl

## ALIGNMENTS

## RESULT 1

US-09-426-072-2  
; Sequence 2, Application US/09426072

; Patent No. 6146869

; GENERAL INFORMATION:

; APPLICANT: Paul Harris

; APPLICANT: Kimberly M. Brown

; TITLE OF INVENTION: Polypeptides Having Phospholipase B

; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same

; FILE REFERENCE: 5951.000-US

; CURRENT APPLICATION NUMBER: US/09/426,072

; CURRENT FILING DATE: 1999-10-21

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 464

; TYPE: PRT

; ORGANISM: Aspergillus oryzae

US-09-426-072-2

Query Match 100.0%; Score 2501; DB 4; Length 464;  
Best Local Similarity 100.0%; Pred. No. 9.7e-235;  
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MKSTALLTGLGLASGLASPVTSYTSVRAPFGYKPGSKESIKLKKVENIWLILLE 60

Db 1 MKSTALLTGLGLASGLASPVTSYTSVRAPFGYKPGSKESIKLKKVENIWLILLE 60

Oy 61 NRSFNDILGGVRRGCLDNPINNPGFCNKNASDPSSGKYCTQAKDYDSVFNDDPHSVTGN 120

Db 61 NRSFNDILGGVRRGCLDNPINNPGFCNKNASDPSSGKYCTQAKDYDSVFNDDPHSVTGN 120

Oy 121 NLEFGYTTYPNNGAIAAGKVVADQSGFLNAQINDYPKLAPEATRVQMGYYTEEEVPTLV 180

Db 121 NLEFGYTTYPNNGAIAAGKVVADQSGFLNAQINDYPKLAPEATRVQMGYYTEEEVPTLV 180

Oy 181 DLVDEFTTFNFWSCVPGPTNPRLCALAGTAAGHKNDNDDFLNTGYSKSIFFAANEKG 240

Db 181 DLVDEFTTFNFWSCVPGPTNPRLCALAGTAAGHKNDNDDFLNTGYSKSIFFAANEKG 240

Oy 241 VSWLNDYDGTNGEFPDPSLFTTYVNTSRSNVNVFNFQDAYLGVLPKFSINPSCCGTN 300

Db 241 VSWLNDYDGTNGEFPDPSLFTTYVNTSRSNVNVFNFQDAYLGVLPKFSINPSCCGTN 300

Oy 301 TNSMHPGTGNSYGEVFKYQIYDAIRGQPDWKTLFLFTYDTGTGFFDHPVPLAVRPDNL 360

Db 301 TNSMHPGTGNSYGEVFKYQIYDAIRGQPDWKTLFLFTYDTGTGFFDHPVPLAVRPDNL 360

Oy 361 TTYTAKNGQKYTLHFDRLGRMPWTWISPYSKKGYIEQYGTDPVTGKDPAPYSATSVLYKT 420

Db 361 TTYTAKNGQKYTLHFDRLGRMPWTWISPYSKKGYIEQYGTDPVTGKDPAPYSATSVLYKT 420

Db	361	TYETAKNGQYTLHFDRLGRMPTWVVISPKSGYIEQYCTDPVTGKPAPVSATSVLKT	420
QY	421	LGYNWDIEFTFRAVHSPSFDHLIGTLREDAPIALTKPTFTSV	464
Db	421	LGYNWDIEFTFRAVHSPSFDHLIGTLREDAPIALTKPTFTSV	464

RESULT 2  
US-08-494-714-2  
; Sequence 2, Application US/08494714  
; Patent NO. 5587290  
; GENERAL INFORMATION:  
; APPLICANT: Klionsky, Daniel  
; APPLICANT: Holzer, Helmut  
; APPLICANT: Desruelle, Monica  
; TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187

```

:
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/494,714
:
:
:

```

[illegible]

```

TELECOMMUNICATION INFORMATION
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-494-714-2

```

Query Match 4.5%; Score 111.5; DB 1; Length 1041;  
Best Local Similarity 19.7%; Pred. NO. 0.083;  
Matches 89; Conservative 64; Mismatches 140; Indels 159; Gaps 24;

QY	22	VTSEY----	TSVREAPFGYKPGSKESLENLKDVENIWLILENRS-----	FDNLTGG	70
Db	446	VDYETHINDVAMASF	SYLNGHEGIDD--EYLRTYTPWKIKNAAGFTYAYVKNSSLGL	503	
QY	71	VRROGLNP-----	INNGPPCNYKNASDFSSGKYCTQAKDYDSVNFDPDHSVTGNLFE	124	
Db	504	YETYNLTDPDFANH	INNGAP--TNAGIKTLLKWTADIGNHLEGVDDKWEISKDI--	558	
QY	125	YCTYTPNNGALASG	KVADQSGFLNAQLNDYKPLAPEARTQV--NGYVTEETPTLVDL	182	
Db	559	---YIPR-----	SSSNITLEYSG-----MNSVEIKQADVTLWYPLCYINDESI--LNN	604	
QY	183	VDFTTNSWFCV--	PGTNPNRILCALAGTAAHG-----	K	217
Db	605	TKDIYYYSERQASG	PAMTYPVFVAAG--LNLHGSSSQSYLYKSVLPYLRAFPAQSEQ	663	
QY	218	NDDDFLNYGISSKSI	--FEAANEK-----GYSWLNDGTNGEFPDLSLFTYYNQTRS	269	

Db	664	SDDNFLTNGLTQAPPFITANGPGLASILFGLTGTRY---	SYEVDPT---KKNLLRRF	717
QY	270	NVPVENFODAYLVGLP-----KFSYNPSCGTNTNSMHPTGNVSGEVFKQYDA	323	
Db	718	N--PIE-----LPLPGGTAIRNPKYWP-----	739	
QY	324	IRQGPQWDKTLFIITYDETTGGFYDH-----VPPFLAVRPDLNLTATETAKNGQYKT	373	
Db	740	-----VLDTIIDHNGTIVHKSGDVP>IHKIPNRSLIHQDINFIYNGSENERKPN	789	
QY	374	LH---FDRLGGRMPTWVISPYSKKGVEQYGT	402	
Db	790	LERSDVRGVDPW-----RMDRYGT	809	

```

RESULT      3
PCT-US96-10782-2
; Sequence 2, Application PC/TUS9610782
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University
; APPLICANT: of California
; TITLE OF INVENTION: SPRESS TOLERANT YEAST MUTANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBBINS, BERLINER & CARSON
; STREET: 201 N. Figueroa Street, 5th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012-2628
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10782
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Berliner, Robert
; REGISTRATION NUMBER: 20,121
; REFERENCE/DOCKET NUMBER: 5555-400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 977-1001
; TELEFAX: (213) 977-1003
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-10782-2

```

Query Match	4.5%;	Score 111.5;	DB 5;	Length 1041;
Best Local Similarity	19.7%;	Pred. NO. 0.083;		
Matches 89;	Conservative 64;	Mismatches 140;	Indels 159;	Gaps 24;

Qy	22	VTSEY----	TSVRAPFGYKPGKSEIENLKDKVENIVMLLENRS-----	FDNITLGG 70
Db	446	VDYEVHINVDVAMASFSYILNGHEGIDD--EYLRVTTWPIIKNAAGFTTAYKYKVNSSLGL 503		
Qy	71	VRRQLNP-----	INNGPCFNKKNASDPSSGKYCTQAKDYDVSFNDPDHSGVTGNNLEF 124	
Db	504	YETYNLTDDPEFANHNGAF---TNAGIKTLKKWATDIGNHGEVVDVPKWEISKDI-- 558		
Qy	125	YGYTPNGAIGSKVADQSGFLNQAOLNDYKPLAPEATRQV--NGYETESEVPTLVLDL 182		
Db	559	---YIPR-----SSNITLEYSG-----MNSSEVIEKQADVTLWYPLGYINDESI--LNNA 604		
Qy	183	VDETTTENSFWSCV-PGPTNPNRLCALAGTAAGHG-----		K 217

Db 605 IKDLYYSEQASGPAMTVPVFAAAG-LLNHGSSQSYLYKSVLPYLRAPAFQSFQ 663  
Qy 218 NDDDELNGISSKI-PEANKE- - - - -GVSWLWYDGTNGEPEPDSLFYVNTSRS 269  
Db 664 SDNEFTNGLTPAPFPLTANGGFLQSILFGLTGRY- - -SYEVDPDF- - -KKINRLR 717  
Qy 270 NVYPVENFODAYLGLVP- - - - -KFSYINPSCGNTNMSHPTGNVSGVFVKQYDA 323  
Db 718 N-PIE- - - - -LPLPGGIAIRNEKVMNP- - - - - 739  
Qy 324 IROGQWOKTLLFTYDEGGFYD- - - - -VPPPLAVRPNLTYTETAKNGOKYT 373  
Db 740 - - - - -VLDIIDHNGTIVHKSGDVPHIKIPNRLIHQDINFNGSENERKEN 789  
Qy 374 LH- - - - -FDRGLGRMPTWISPSKKGYIEQYGT 402  
Db 790 LERRDVRGDP- - - - -RMDRYGT 809  
RESULT 4  
5223424-13  
; Patent No. 5223424  
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,  
; RICHARD D.  
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND  
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO  
; ACID SEQUENCE  
; NUMBER OF SEQUENCES: 16  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/225,032  
; FILING DATE: 27-JUL-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 78,519  
; FILING DATE: 27-JUL-1987  
; APPLICATION NUMBER: 933,107  
; FILING DATE: 20-NOV-1986  
; APPLICATION NUMBER: 902,887  
; FILING DATE: 02-SEP-1986  
; APPLICATION NUMBER: 887,140  
; FILING DATE: 17-JUL-1986  
; APPLICATION NUMBER: 823,102  
; FILING DATE: 27-JAN-1986  
; APPLICATION NUMBER: 773,430  
; FILING DATE: 06-SEP-1985  
; SEQ ID NO:13:  
; LENGTH: 579  
5223424-13  
Query Match 4.4%; Score 109.5; DB 6; Length 579;  
Best Local Similarity 21.4%; Pred. No. 0.051;  
Matches 86; Conservative 44; Mismatches 151; Indels 121; Gaps 22;  
Qy 110 FNDDHSVTCNNL- - - - -EFY- - - - -GTYP- - - - -NNGAISGKVVADQS 145  
Db 121 FNPADWOLISNMTINLVSEQEIENNVVLKTTESATSPSKIYNNDLTLASLVALDTN 180  
Qy 146 GFLMAQLNDYKLAPEEATRQVMGYTTEEVPTLVLDVDEFTTFSWFSVCVPGPNRNL 205  
Db 181 NTL- - - - -PYTPAARSET- - - - -LGFY- - - - -PWLPT- - - - -KPTQRYLYLSCIRNLNPP- - - 221  
Qy 206 CALAGTAAGHKNDODFLNGYSSKSIIFEAANEKGVSWLNYDG- - -TNG- - -EFEPDSLFF 260  
Db 222 - - - - -TYSQSQITDSIQGLHSDIMFTIENAVP IHLRTGDEFSTGIYHFDKPKL 276  
Qy 261 TYVNOTSRNVVVENFEQ- - - - -DAYLGLVLPKFSYINPSCGNT- - -NSMHTPGNVSYGE 314  
Db 277 THSNQTNRLSLGLPKVLPTEPTTEGQDHPGTL- - - - -GANTKRGYHQITNNSTE 326  
Qy 315 VFVKQYDAIROGP-QWDKTLLFTYDEGGFYDVPPLAVRPNLTYTETAKNGOKYT 373  
Db 327 A- - - - -TALRPAQGYNTPYVNNFDYNSGGPFL- - - - -TPVPTADTYTDEPENGAIREF 376

Qy 374 L- - - - -HFDRGLGRMPTWISPSKKGYI- - -EQVG- - - - -TDPVTGKP 409  
Db 377 MGYQHGLTTSQELERYTFNPSKCGRAPKQFNOQAPLNLENTNNGTLLPSDPIGGKS 436  
Qy 410 - - - - -APYSA- - - - -TSVLKTLGYLWDIE- - -DFTPRV 434  
Db 437 NKHEWNTLTYGTLTALNTPAVFPNGQIMDKELDTDLAPRL 478  
RESULT 5  
US-08-049-783-2  
; Sequence 2, Application US/08049783  
; Patent No. 5439881  
; GENERAL INFORMATION:  
; APPLICANT: Narva, Kenneth E  
; APPLICANT: Schwab, George E  
; APPLICANT: Payne, Jewel M  
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active  
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jeff Lloyd  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/049,783  
; FILING DATE: 19930419  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lloyd, Jeff  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 2 (PS33F2):  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1257 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; INDIVIDUAL ISOLATE: PS33F2  
; IMMEDIATE SOURCE:  
; CLONE: PS33F2a  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1257  
US-08-049-783-2  
Query Match 4.3%; Score 107; DB 1; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.31;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
Qy 6 LLTGLGLLASLG-LASPVTSYTSVREAPFGKPGSKESINLKDKNVWVLLLEN- - - 61  
Db 74 VLSVLTLPVGTAVSAASTIVSPKIFGDKPNKAKNIFEELKPOEALIQODITNQD 133  
Qy 62 - - - - -RSFDNLGVRRQGLDNPINNGFCNKNASDPSSGKYCTQAKDYDSVFN- - - 112  
Db 134 AINOKFDSLQKTINLYTV- - - - -AIDNDYVTAK- - - - -TQLENLSILTSISIF 180

QY 113 -PDHVSVTGNLEFYGTYPNNGAIAAGKVV-ADQSGFLNAQLNDYPKLAPEATROVMGY 170  
Db 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234  
QY 171 YTEEVPTLVDLVEFTTF--NSWFSVCPGTPNRLCALAGTAAGHCKNDODFLNKGIS 228  
Db 235 HTEAVTKAFLGKFLSDVNSY-----NKKANY--- 264  
QY 229 SKSIFEAANEKGVSNLYDGTNGEPEPSLFTYVN-----QTSRNNVYPVENF 277  
Db 265 IKGTEWLDLVALWPTFDPDHYQKEVELEFTRTISPIYQVPKMNQNTSSIVPSDLF 324  
QY 278 FQDAYLGLVLPK--FS-----YINPSCCGT-----NTNSMHPGTGN 309  
Db 325 H---YOGDLVKLEFSTRDNDGLAKIFTGIRNTFYKSPNTHETIYHVDYSYNTQS---SGN 378  
QY 310 VSYG-----EVEVKQIYD-----AIROGPQMDKTLF 336  
Db 379 ISRGSSNPIDLNPIIISTCIRNSFYKAIAAGSSVLNFKDGTQGYAFQAAP----- 430  
QY 337 ITYDETGGEYD--VPPPLAVRPDNLTYETAKNG-----QKYLHFDRLGGRMPTWVI 388  
Db 431 -----TGGAWDSHFTESDGAPEGHKLNIIYTSFGDTRDFINVYTLISTPTINELSTEKI 485  
QY 389 SPY-SKKGYIEQYGDPTVTKP 409  
Db 486 KGFPAAEKGYIKNOGIMKYGKP 507

## RESULT 6

US-08-158-232-6

; Sequence 6, Application US/08158232

; Patent No. 5596071

; GENERAL INFORMATION:

; APPLICANT: Payne, Jewel

; APPLICANT: Kennedy, M. Keith

; APPLICANT: Randall, John Brooks

; APPLICANT: Meier, Henry

; APPLICANT: Uick, Heidi Jane

; APPLICANT: Foncarrada, Luis

; APPLICANT: Schnepf, H. Ernest

; APPLICANT: Schwab, George E.

; APPLICANT: Fu, Jenny

; TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active

; CORRESPONDENCE ADDRESS:

; ADDRESS: David R. Saliwanchik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/158,232

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/887,980

; FILING DATE: 22-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/797,645

; FILING DATE: 25-NOV-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/703,977

; FILING DATE: 22-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; APPLICANT: Payne, Jewel M.

NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCU104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-158-232-6

Query Match 4.3%; Score 107; DB 1; Length 1257;

Best Local Similarity 19.3%; Pred. No. 0.31;

Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LITGLGLLASLG-LASPTSEVTSVREAPFGYKPKGSKESIENLKDKVENIVMLLEN--- 61

Db 74 VLSVLTLPYEGTVASAASITVFWPKIFGDKFNKAKNIFEELKQIEALIQDITNYQD 133

QY 62 ----RSFNILGGVRRGCLDNPINNGPCNKNASDPSSGKYCTQAKDYDSVND----- 112

Db 134 AINOKKFDLSQKTIINLYTV--AIDNNXYTAK-----TOLENLNLSILFSDISIF 180

QY 113 -PDHVSVTGNLEFYGTYPNNGAIAAGKVV-ADQSGFLNAQLNDYPKLAPEATROVMGY 170

Db 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234

QY 171 YTEEVPTLVDLVEFTTF--NSWFSVCPGTPNRLCALAGTAAGHCKNDODFLNKGIS 228

Db 235 HTEAVTKAFLGKFLSDVNSY-----NKKANY--- 264

QY 229 SKSIFEAANEKGVSNLYDGTNGEPEPSLFTYVN-----QTSRNNVYPVENF 277

Db 265 IKGTEWLDLVALWPTFDPDHYQKEVELEFTRTISPIYQVPKMNQNTSSIVPSDLF 324

QY 278 FQDAYLGLVLPK--FS-----YINPSCCGT-----NTNSMHPGTGN 309

Db 325 H---YOGDLVKLEFSTRDNDGLAKIFTGIRNTFYKSPNTHETIYHVDYSYNTQS---SGN 378

QY 310 VSYG-----EVEVKQIYD-----AIROGPQMDKTLF 336

Db 379 ISRGSSNPIDLNPIIISTCIRNSFYKAIAAGSSVLNFKDGTQGYAFQAAP----- 430

QY 337 ITYDETGGEYD--VPPPLAVRPDNLTYETAKNG-----QKYLHFDRLGGRMPTWVI 388

Db 431 -----TGGAWDSHFTESDGAPEGHKLNIIYTSFGDTRDFINVYTLISTPTINELSTEKI 485

QY 389 SPY-SKKGYIEQYGDPTVTKP 409

Db 486 KGFPAAEKGYIKNOGIMKYGKP 507

## RESULT 7

US-08-304-626-6

; Sequence 6, Application US/08304626

; Patent No. 5616495

; GENERAL INFORMATION:

; APPLICANT: Payne, Jewel M.

APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates  
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding  
TITLE OF INVENTION: Hymenopteran-Active Toxins  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,626  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/887,980  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ 104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-304-626-6

Query Match 4.3%; Score 107; DB 1; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.31;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
QY 6 LLTGIGLALSLG-LASPVTSYTSVREAPGYKPGKSTENLKKVENIWLLEN--- 61  
Db 74 VLSVLITPEVTGASAASTIVSFIPKIFGDKPNAKNIPEELKPGQLEALIQDITNQD 133  
QY 62 ----RSPDNILGVRROGLDNPINGPFCNKNASDPSSGKYCTQAKDYDSVND----- 112  
Db 134 AINOKKFDLSQKTLNLYTV--AIDNDYVTAK-----TOLNLSILTSDISIF 180  
QY 113 -PDHSVTGNLLEFYCTYPNNGATASGVV-ADOSGFLNQLNDYDKLAPEATQVMGY 170  
Db 181 IPEGYETG-GLPYAWANAHILLRLDAIVNAEKLGFSDKEVDTHRY-----IKMTIH 234

QY 171 YTEEEVPTLVLDVDEFTTF--NSWFSQVPGPTNPRLCALAGTAGHGKNDODFLNYGIS 228  
Db 235 HTEAVIKAFNLGDKFKSLDVSNS- - - - -NKKANY- - - - - 264  
QY 229 SKSIFENANEKGVSWLNYDGTNGEPEPDSLFTTYN- - - - -QTSRSNVVYVNF 277  
Db 265 IKGTEMVLDLVALMPTFDPDHYOKEVEIEFTTRTSSPIYQVPKMNQMTSSSIVPSDLF 324  
QY 278 FQDAYLGVLPK--FS- - - - -YINPSCCGT- - - - -NTNSMHPGTN 309  
Db 325 H--YQGLVKLEFSTRTDNDGLAKIFTGIRNTFYKSPNTHETVYHVDYSYNTQS- - - - -SGN 378  
QY 310 VSYG- - - - -EVEFKQIYD- - - - -AIQGPQWDKTLF 336  
Db 379 ISRSSNPIDLNPNPISTCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP- - - - - 430  
QY 337 ITYDEGTGYD- - - - -VPPPLAVRPDNLTYTETAKNG- - - - -QKYLHFDRLGGRMPTWVI 388  
Db 431 ----TGGAWDSFTESDGAPEGHKLNLYTTSPTGLRDFINVTTLISTPTINELSTEKI 485  
QY 389 SPY-SKKGYIEQYGFDPVTKGP 409  
Db 486 KGPPAEKGIKQNGIMKYKGP 507  
RESULT 8  
US-08-316-301A-6  
Sequence 6, Application US/08316301A  
Patent No. 5753492  
GENERAL INFORMATION:  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Foncerrada, Luis  
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
TITLE OF INVENTION: Which Code Therefor  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,301A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/871,510  
FILING DATE: 23-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/693,018  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/565,544  
FILING DATE: 10-AUG-1990  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/084,653  
FILING DATE: 12-AUG-1987  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/830,050  
FILING DATE: 31-JAN-1992

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
REGISTRATION NUMBER: 35,589  
REFERENCE/DOCKET NUMBER: MA20CCCD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6 (PS33F2):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC 2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-316-301A-6

Query Match 4.3%; Score 107; DB 1; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.31;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
QY 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESIKENLKDVENIWLILEN--- 61  
DB 74 VLSVLTLPVETVSAASTIVSFIPWFKIFGDKPNAKNIFEELPKQIEALIQODITNYQD 133  
QY 62 ---RSPDNLGVRQCLDNPNGPCNKNASDPSSGKYCTQAKDYDSVEND--- 112  
DB 134 AINOKKEDSLQKTLNLYTV--AIDNDYVTK-----TQLENLSILTSISIF 180  
QY 113 -PDHSVTGNLLEFYCTYPNNGAIASGVV-ADQSGFLNQLNDYKLAPEATPQVNGY 170  
DB 181 IPEGYETG-GLPYAMAVANAHILLRDAIVNAEKLGSFSDKEVDTHKKY-----IKMTIH 234  
QY 171 YTEEVPTLVDLDEFTTF--NSWFSVCPPTNPRLCALAGTAGHGKNDDEFLNYGIS 228  
DB 235 HTEAVIKAFGLNGDKFKSLDVNSY-----NKKANY--- 264  
QY 229 SKSIFEANEKGVSWLNDCTNGEFPDLSLFTYVN-----QTSRSNVVPYENF 277  
DB 265 IKGTEMLVDLVALMPTFDPHYOKEVEIEFTRTISSPIQVPKKNQNTSSIVPSDLF 324  
QY 278 FQDAYLGLVLPK--FS-----YINPSCCGT-----NTNSMHPITGN 309  
DB 325 H---YQGLVKLEFSTRDNDGLAKIFTGIRNTEYKSPNTHYHVDVSYNTQS---SGN 378  
QY 310 VSYG-----EVFKQIYD-----AIRGQFQWDKTLFF 336  
DB 379 IYDGTGGFYDH--VPPPLAVRPNLTYTETAKNG-----QKTYLHEDRLGCRMPTWVI 388  
DB 431 -----TGGANDHSFIESDCAPEGHKLNYIYTSFGDTRLDFINVTLISTPTINELSTEKI 485  
QY 389 SPY-SKKGYIEQGTDPVTGKP 409  
DB 486 KGFFPAEKGYIKNGIMKYGRP 507

## RESULT 9

US-08-611-928-6  
; Sequence 6, Application US/08611928  
; Patent No. 5824792

GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,928  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/158,232  
FILING DATE: 24-NOV-1993  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-611-928-6

Query Match 4.3%; Score 107; DB 2; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.31;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;



```

Qy 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESITENLKOKVENIWMILLEN--- 61
Db 74 VLSVLTLPVEGTVASASTIVSFVWPKIFGDKPNAKNIPEELKPOIEALIQOQDITNYOD 133
Qy 62 ----RSFDNLLGVRROGLDNPINNGPFCNKYKNASDPSSGKYCTQAKDYDSVFN----- 112
Db 134 AINQKFKDSLOKTNLYTV--AIDNNDVYTK-----TOLENLSILTSISIF 180
Qy 113 -PDHVSVTGNLEFYGTTPNNGAIAAGKVV-ADQSGFLNAQLNDYKPLAPEATQVMGY 170
Db 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGSDEKVEDTHKKY-----IKMTIHN 234
Qy 171 YTEEEVPTLVLDVDEFTF--NSWFSVCPGPTNPRLCALAGTAAGHGKNDODFLNYGIS 228
Db 235 HTEAVTKAFLAGLDFKSLDVSY-----NKKANY--- 264
Qy 229 SKSIFPAANEKGVSNLNYDGTNGEPEPDSLFFTVN-----QTSRSNVVVFENF 277
Db 265 IKGTEWVLDLVALWPTDFPDHYQKEVEIEFTRTISSPIYQVPVKNMONTSSIVPSDLF 324
Qy 278 FODAYLGLVLPK--FS-----YINPSCGT-----NTNSMHPGTG 309
Db 325 H----YQGLDLVKLEFSTFDNDGLAKIFGIRNTFYKSPNTHETVHVSFYNTQS---SGN 378
Qy 310 VSYG-----EVFVKQIYD-----AIROGPQWDKTLF 336
Db 379 ISRGSSNPITIDLANNPITCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP----- 430
Qy 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLTHFDRLGGRMPTWVI 388
Db 431 ----TGGAWDHSFIESDGAPEGHKLNIIYSPGDTLRDFINVTYTLISTPTINELSTEKI 485
Qy 389 SPY-SKKGYIEQYGTDPVTGKP 409
Db 486 KGPPAEKGYIKNOGIMYYGKP 507

```

## RESULT 10

```

US-09-173-891-6
; Sequence 6, Application US/09173891
; Patent No. 6077937
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Kennedy, M. Keith
; APPLICANT: Randall, John Brooks
; APPLICANT: Meier, Henry
; APPLICANT: Uick, Heidi Jane
; APPLICANT: Fonceerrada, Luis
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Schwab, George E.
; APPLICANT: Fu, Jenny
; TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active
; AGAINST HYMENOPTERAN PESTS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/173,891
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/158,232

```

```

; FILING DATE:
; APPLICATION NUMBER: US 07/887,980
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/797,645
; FILING DATE: 25-NOV-1991
; PRIOR APPLICATION DATA: US 07/703,977
; FILING DATE: 22-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/SCJ104.C1
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS33F2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC2316) B-18785
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1257
; US-09-173-891-6

```

Query Match 4.3%; Score 107; DB 3; Length 1257;

Best Local Similarity 19.3%; Pred. No. 0.31;

Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

```

Qy 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESITENLKOKVENIWMILLEN--- 61
Db 74 VLSVLTLPVEGTVASASTIVSFVWPKIFGDKPNAKNIPEELKPOIEALIQOQDITNYOD 133
Qy 62 ----RSFDNLLGVRROGLDNPINNGPFCNKYKNASDPSSGKYCTQAKDYDSVFN----- 112
Db 134 AINQKFKDSLOKTNLYTV--AIDNNDVYTK-----TOLENLSILTSISIF 180
Qy 113 -PDHVSVTGNLEFYGTTPNNGAIAAGKVV-ADQSGFLNAQLNDYKPLAPEATQVMGY 170
Db 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGSDEKVEDTHKKY-----IKMTIHN 234
Qy 171 YTEEEVPTLVLDVDEFTF--NSWFSVCPGPTNPRLCALAGTAAGHGKNDODFLNYGIS 228
Db 235 HTEAVTKAFLAGLDFKSLDVSY-----NKKANY--- 264
Qy 229 SKSIFPAANEKGVSNLNYDGTNGEPEPDSLFFTVN-----QTSRSNVVVFENF 277
Db 265 IKGTEWVLDLVALWPTDFPDHYQKEVEIEFTRTISSPIYQVPVKNMONTSSIVPSDLF 324
Qy 278 FODAYLGLVLPK--FS-----YINPSCGT-----NTNSMHPGTG 309
Db 325 H----YQGLDLVKLEFSTFDNDGLAKIFGIRNTFYKSPNTHETVHVSFYNTQS---SGN 378
Qy 310 VSYG-----EVFVKQIYD-----AIROGPQWDKTLF 336
Db 379 ISRGSSNPITIDLANNPITCIRNSFYKAIAGSSVLVNFKDGTOGYAFAQAP----- 430
Qy 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLTHFDRLGGRMPTWVI 388
Db 431 ----TGGAWDHSFIESDGAPEGHKLNIIYSPGDTLRDFINVTYTLISTPTINELSTEKI 485
Qy 389 SPY-SKKGYIEQYGTDPVTGKP 409

```

```

; Sequence 6, Application PC/TUS9203624
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schnepf, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Foncecrada, Luis
; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes
; TITLE OF INVENTION: Which Code Therefor
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/03624
; FILING DATE: 19920501
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA20C2C1C1
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6 (PS33F2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS33F2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC 2316) B-18785
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1257
; PCT-US92-03624-6

Query Match 4.3%; Score 107; DB 4; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPKSKESIEENLKDKVENIVWLILEN--- 61
DB 74 VLSVLTLPVEGTVASASTVSIWPKIFGDKPNAKNIFEELKPKQIEALIQDITVQD 133
QY 62 ----RSFDNIGVRRQGLDNPINNGPFCNKNASDPSSGKYCTQAKYDSVFN----- 112
DB 134 AINQKRFDSLQKTINLYTV--AIDNNDVYVTA-----TOLNLSILTSDISIF 180
QY 113 -PDHSVTGNLFEYGYTPNNGAIGSKV-ADQSGFLNAQLNDYKPLAPEATRVQVGY 170
DB 181 IPEGYETG-GLPYAYAMANAHIILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234
QY 171 YTEEVPITLVLDVDEFTTF--NSWFCVPGPTNPNRLCALAGTAAGHKNDDEFINYGIS 228
DB 235 HTEAVIKAFGLGDKFKSLDVNSY-----NKKANY--- 264

RESULT 11
US-09-076-137-6
; Sequence 6, Application US/09076137B
; Patent No. 6166195
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schnepf, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Foncecrada, Luis
; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes Which Code
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MA-20CCCD2
; CURRENT APPLICATION NUMBER: US/09/076.137B
; CURRENT FILING DATE: 1998-05-12
; EARLIER APPLICATION NUMBER: 08/316,301
; EARLIER FILING DATE: 1994-09-30
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1257
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
; US-09-076-137-6

Query Match 4.3%; Score 107; DB 4; Length 1257;
Best Local Similarity 19.3%; Pred. No. 0.31;
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPKSKESIEENLKDKVENIVWLILEN--- 61
DB 74 VLSVLTLPVEGTVASASTVSIWPKIFGDKPNAKNIFEELKPKQIEALIQDITVQD 133
QY 62 ----RSFDNIGVRRQGLDNPINNGPFCNKNASDPSSGKYCTQAKYDSVFN----- 112
DB 134 AINQKRFDSLQKTINLYTV--AIDNNDVYVTA-----TOLNLSILTSDISIF 180
QY 113 -PDHSVTGNLFEYGYTPNNGAIGSKV-ADQSGFLNAQLNDYKPLAPEATRVQVGY 170
DB 181 IPEGYETG-GLPYAYAMANAHIILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234
QY 171 YTEEVPITLVLDVDEFTTF--NSWFCVPGPTNPNRLCALAGTAAGHKNDDEFINYGIS 228
DB 235 HTEAVIKAFGLGDKFKSLDVNSY-----NKKANY--- 264

QY 229 SKSIFEAAEKGVSWLNDGTNGEFPDPSLFFTYVN-----QTSRNVVPVNEF 277
DB 265 IKGTEVMDLVALWPTDPDHYQKEVEIEFTRTISSPIYQVPKKNQNTSSSIYPSDLF 324
QY 278 FQDAYLGLPK-FS-----VINPSCCGI-----NTNSMHPGTGN 309
DB 325 H---YQGLLVKLEFSTRDNDGLAKIFTGIRNFTYKSPNTHYHVDFTSYNQF---SGN 378
QY 310 VSYG-----EVFYKQIYD-----AIRQGPQMDKTLFF 336
DB 379 ISRGSSNPIDLNPIITSTCIRNSFYKAIAGSSVLNVFKDGTQGYAFAQAP----- 430
QY 337 ITYDETGGFYDH--VPPPLAYRPNLTYTETAKNG-----QKYLHEDLGRGPTWVI 388
DB 431 -----TGGAWDHSFTESGAGEGKLMNYIYTPGDTLRDFINWYTLISTPTINELSTEKI 485
QY 389 SPY-SKKGYIEQYGDPTVTKP 409
DB 486 KGFPAAEKGYIKNOGIMKYGKP 507

RESULT 12
PCT-US92-03624-6
```



```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 737 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-336-308A-4

Query Match      4.2%; Score 104.5; DB 3; Length 737;
Best Local Similarity 20.0%; Pred. No. 0.23;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

QY 36 YKPGKSESTENLKDVENIWMILLENRSPDNLGVRQGLDNPINNGPFCNYKNASDPS 95
DB 223 YEFGRTPVE---EKQGRMIVIVAKYE---GDIK-----DFVDMKQRG-- 262
QY 96 SGKYCTQAKDYSDVNDHSVGTGNLLEFY--GYTPNNGAIASGVKVDASGFLNAQLN 153
DB 263 ---LRTEVK---VAEDIASPVTAIAIOQFVKQYEKEGNDLTYVLLVGDH-----K 307
QY 154 DYP-KLAPEEATQVMG-----YYTE-----EEVPTLDLVDEF-----TTFNS 191
DB 308 DIPAKITPGIKSDQYGGIYGVNDHYNEVFIGRFSCEKEDLKTQIDRTIHYERNITTEDK 367
QY 192 WFCVCPGPNPRLC---ALAGTAAGHKNDLDFNY-----GISS 229
DB 368 WLG-----QALCIASAGGSPADNGESDIQHENVIANLLTQYTKIKCYDPGVTP 419
QY 230 KSIFEAAKEGYSWLNVDGTNGEPEPDSLFF--TYVNOTSRSNVVPVENFFODA----- 281
DB 420 KNIIDAFN--GGISLVNYTGHGSETAMGTSHFGTTHVKQLTNSQLP---FIFDVACVNGD 475
QY 282 YLGVLPRF-----SYNPS-----CGTNTNSMHP 306
DB 476 FLFSMPCFAELMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNKR 535
QY 307 T-GNVSIGEYFV---KQYDAIROGPQW---DKTLFIITYDETFGGFYDHPVPPPLAY-RP 357
DB 536 TFGGVTMNGMFAMVEYKKDGEKMLDTWTVFGDPSLLVRTL-----VPTKMQVTAP 586
QY 358 DNLTYTET-----AKNGQY-----TLHFDRLGGRMP-TWVLSPY 391
DB 587 AQINLTDASVNVSCDYNCAIATISANGKMFSGAVVNGTATINLTGLTNESTLTLTVVGY 646
QY 392 SKKGYIQOYTDPTGPKAPY-----SATSVLTKLGLYMD 426
DB 647 NKETVIKTINTN---GEPNPYPVSNLTATTGQKVKTLKWD 684

RESULT 15
US-08-822-324-4
; Sequence 4, Application US/08822324
; Patent No. 6129917
; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

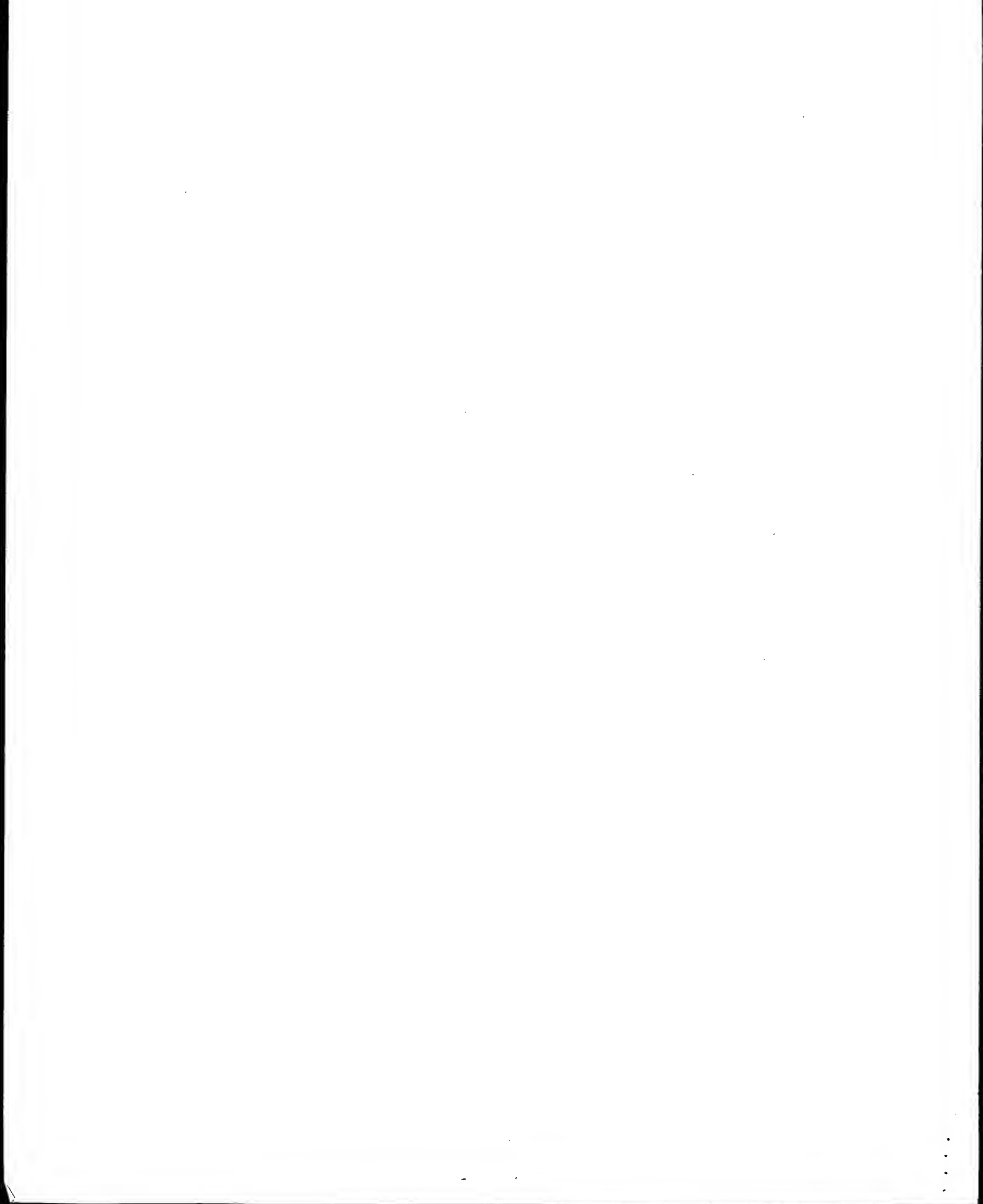
```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 488-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 737 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-822-324-4

Query Match      4.2%; Score 104.5; DB 3; Length 737;
Best Local Similarity 20.0%; Pred. No. 0.23;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

QY 36 YKPGKSESTENLKDVENIWMILLENRSPDNLGVRQGLDNPINNGPFCNYKNASDPS 95
DB 223 YEFGRTPVE---EKQGRMIVIVAKYE---GDIK-----DFVDMKQRG-- 262
QY 96 SGKYCTQAKDYSDVNDHSVGTGNLLEFY--GYTPNNGAIASGVKVDASGFLNAQLN 153
DB 263 ---LRTEVK---VAEDIASPVTAIAIOQFVKQYEKEGNDLTYVLLVGDH-----K 307
QY 154 DYP-KLAPEEATQVMG-----YYTE-----EEVPTLDLVDEF-----TTFNS 191
DB 308 DIPAKITPGIKSDQYGGIYGVNDHYNEVFIGRFSCEKEDLKTQIDRTIHYERNITTEDK 367
QY 192 WFCVCPGPNPRLC---ALAGTAAGHKNDLDFNY-----GISS 229
DB 368 WLG-----QALCIASAGGSPADNGESDIQHENVIANLLTQYTKIKCYDPGVTP 419
QY 230 KSIFEAAKEGYSWLNVDGTNGEPEPDSLFF--TYVNOTSRSNVVPVENFFODA----- 281
DB 420 KNIIDAFN--GGISLVNYTGHGSETAMGTSHFGTTHVKQLTNSQLP---FIFDVACVNGD 475
QY 282 YLGVLPRF-----SYNPS-----CGTNTNSMHP 306
DB 476 FLFSMPCFAELMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNKR 535
QY 307 T-GNVSIGEYFV---KQYDAIROGPQW---DKTLFIITYDETFGGFYDHPVPPPLAY-RP 357
DB 536 TFGGVTMNGMFAMVEYKKDGEKMLDTWTVFGDPSLLVRTL-----VPTKMQVTAP 586
QY 358 DNLTYTET-----AKNGQY-----TLHFDRLGGRMP-TWVLSPY 391
DB 587 AQINLTDASVNVSCDYNCAIATISANGKMFSGAVVNGTATINLTGLTNESTLTLTVVGY 646
QY 392 SKKGYIQOYTDPTGPKAPY-----SATSVLTKLGLYMD 426
DB 647 NKETVIKTINTN---GEPNPYPVSNLTATTGQKVKTLKWD 684

Search completed: December 4, 2001, 15:08:54
Job time: 114 sec
```











```
; SEQ ID NO 4
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-426-072-4

Query Match      0.9%; Score 21; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2030 ccataccttttcggtataagt 2050
    |||||
Db 30 CCATACCTTTTCGGTATAAGT 10

RESULT 3
US-08-814-095-7/c
; Sequence 7, Application US/08814095
; Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6026183thwestern Highway, Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,095
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Montgomery, Ilene N.
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Cosmid including ACHE gene"
; DESCRIPTION: promotor, ACHE gene and ARS gene"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 7q22
; FEATURE:
; NAME/KEY: promotor
; LOCATION: 4089..22464
; OTHER INFORMATION: /function= "ACHE Promotor"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 22465..22537
; OTHER INFORMATION: /standard_name= "ACHE Promotor"

;
; OTHER INFORMATION: /function= "non-translated"
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 24090..25177 experimental
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /function= "(translation start:
; OTHER INFORMATION: 24110)"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25524..26009 experimental
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27005..27274 experimental
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27255..28007 experimental
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 27385..27387
; FEATURE:
; NAME/KEY: exon
; LOCATION: 28008..28129 experimental
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 28129..28131
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (34528..34895)
; OTHER INFORMATION: /function= "arsenite resistance
; OTHER INFORMATION: gene"
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (34092..34358)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33779..33963)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33493..33591)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33297..33408)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 5
```

```
FEATURE:
NAME/KEY: exon
LOCATION: complement (32959..33094)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: exon
LOCATION: complement (32569..32628)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 7
FEATURE:
NAME/KEY: exon
LOCATION: complement (32386..32468)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 8
FEATURE:
NAME/KEY: exon
LOCATION: complement (31894..32080)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 9
FEATURE:
NAME/KEY: exon
LOCATION: complement (31363..31534)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 10
FEATURE:
NAME/KEY: exon
LOCATION: complement (31131..31284)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 11
FEATURE:
NAME/KEY: exon
LOCATION: complement (30816..31011)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 12
FEATURE:
NAME/KEY: exon
LOCATION: complement (30470..30626)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 13
FEATURE:
NAME/KEY: exon
LOCATION: complement (30187..30274)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 14
FEATURE:
NAME/KEY: exon
LOCATION: complement (29945..30073)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 15
FEATURE:
NAME/KEY: exon
LOCATION: complement (29664..29856)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number= 16
US-08-814-095-7
```

```
Query Match 0.8%; Score 19; DB 3; Length 35060;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 536 gtctggcctcttgccctc 554
Db 6964 GTCTGGCCTCTTGGCCTC 6946
```

```
RESULT 4
US-09-426-072-3
; Sequence 3, Application US/09426072
; Patent No. 6146869
; GENERAL INFORMATION:
; APPLICANT: Paul Harris
```

```
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Polypeptides Having Phospholipase B
; FILE REFERENCE: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5951.000-US
; CURRENT APPLICATION NUMBER: US/09/426.072
; CURRENT FILING DATE: 1999-10-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-426-072-3
```

```
Query Match 0.8%; Score 18; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 511 atgaagtcaccgcgtctg 528
Db 10 atgaagtcaccgcgtctg 27
```

```
RESULT 5
US-09-058-489-25
; Sequence 25, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahm, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: the Y Chromosome
; FILE REFERENCE: WHI97-08pa
; CURRENT APPLICATION NUMBER: US/09/058.489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041.877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-25
```

```
Query Match 0.7%; Score 17; DB 3; Length 556;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 808 gaggcgtgcgcgcgcacaa 824
Db 214 gaggcgtgcgcgcgcacaa 230
```

```
RESULT 6
US-08-265-086-1
; Sequence 1, Application US/08265086
; Patent No. 5576191
; GENERAL INFORMATION:
; APPLICANT: Gayle, Margit
; APPLICANT: Slack, Jennifer
; APPLICANT: Gross, Hans-Juergen
; APPLICANT: Sims, John E.
; APPLICANT: Dower, Steven K.
; TITLE OF INVENTION: No. 5576191el Cytokine That Binds ST2
; CORRESPONDENCE ADDRESS: 5
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
```

STATE: Washington  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/265,086  
FILING DATE: June 17, 1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2825  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: huST2-11g  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 88..771  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 88..156  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 157..768  
US-08-265-086-1

Query Match 0.7% Score 17; DB 1; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 630 caaggagtcattgaga 646  
|||||  
DB 549 CAAGGAGTCATTGAGA 565

RESULT 7  
US-09-064-693A-19  
; Sequence 19, Application US/09064693A  
; Patent No. 6210937  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Thomas E.  
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION  
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Gary Goodson  
; ADDRESSEE: INEEL--Lockheed Martin Idaho  
; ADDRESSEE: Technologies Co.  
; STREET: P.O. Box 1625  
; CITY: Idaho Falls  
; STATE: Idaho  
; COUNTRY: USA  
; ZIP: 83415-3810  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/265,086  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: W. Gary Goodson  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2825  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822

COMPUTER: Toshiba Satellite Pro T2150CDS  
OPERATING SYSTEM: Windows95  
SOFTWARE: Word Perfect 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/064,693A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: W. Gary Goodson  
REGISTRATION NUMBER: 22,387  
REFERENCE/DOCKET NUMBER: LIT-PI-296  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (208) 526-9469  
TELEFAX: (208) 526-8339  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2456 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-064-693A-19

Query Match 0.7% Score 17; DB 4; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2078 aaagtgtatgacgt 2094  
|||||  
DB 1511 AAAGTGTATGACGT 1527

RESULT 8  
US-09-064-693A-25  
; Sequence 25, Application US/09064693A  
; Patent No. 6210937  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Thomas E.  
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION  
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Gary Goodson  
; ADDRESSEE: INEEL--Lockheed Martin Idaho  
; ADDRESSEE: Technologies Co.  
; STREET: P.O. Box 1625  
; CITY: Idaho Falls  
; STATE: Idaho  
; COUNTRY: USA  
; ZIP: 83415-3810  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Word Perfect 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/064,693A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: W. Gary Goodson  
; REGISTRATION NUMBER: 22,387  
; REFERENCE/DOCKET NUMBER: LIT-PI-296  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (208) 526-9469  
; TELEFAX: (208) 526-8339

; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-25

Query Match 0.7%; Score 17; DB 4; Length 6641;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2078 aaagtcatgtaacgt 2094
Db 5696 AAAGTATGATGACGT 5712

RESULT 9
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PROCESSOR (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,901C
; FILING DATE: MAY 23, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/057,531
; FILING DATE: APRIL 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BOGOSIAN, ELIZABETH A
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/273
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 OR 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50341
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: L5 shuttle plasmid sequence
; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: L5 mycobacteriophage
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION: No. 5750384e
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; PAGES:
; DATE:

```

; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1

Query Match          0.7%  Score 17; DB 1; Length 50341;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
|||||
Db 20287 TCTGGGACATCGAGGAC 20303

RESULT 11
US-09-075-904-1
; Sequence 1, Application US/09075904
; Patent No. 5994137
; GENERAL INFORMATION:
; APPLICANT: Jacobs, et al.
; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075,904
; FILING DATE: May 11, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/247,901
; FILING DATE: May 23, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Elizabeth A
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/475
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50341
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: L5 shuttle phasmid sequence
; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: L5 mycobacteriophage
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:

; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION: No. 5994137e
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1

Query Match          0.7%  Score 17; DB 2; Length 50341;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
|||||
Db 20287 TCTGGGACATCGAGGAC 20303

RESULT 12
US-09-426-436-1
; Sequence 1, Application US/09426436
; Patent No. 6225066
; GENERAL INFORMATION:
; APPLICANT: William R. Jacobs, Jr.
; APPLICANT: Barry R. Bloom
; APPLICANT: Graham F. Hatfull
; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/426,436
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,557
; FILING DATE:
; FILING DATE:
; APPLICATION NUMBER: US/08/057,531
; APPLICATION NUMBER: 07/833,431
; FILING DATE: February 7, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: 96700/238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 52297
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: phage genome sequence
DESCRIPTION: phage genome sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: not applicable.
ORIGINAL SOURCE:
ORGANISM: mycobacteriophage L5
STRAIN: not applicable
DEVELOPMENTAL STAGE: not applicable
HAPLOTYPE: not applicable
TISSUE TYPE: not applicable
CELL LINE: not applicable
ORGANELLE: not applicable
IMMEDIATE SOURCE: mycobacteriophage L5 particles
POSITION IN GENOME: entire genome
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Hatfull and Sarkis
TITLE: DNA Sequence, Structure and Gene
TITLE: Expression of Mycobacteriophage L5:
TITLE: A Phage System for Mycobacterial
TITLE: Genetics
JOURNAL: Molecular Microbiology
VOLUME: 7
PAGES: 395-405
DATE: 1993
US-08-705-557-1

Query Match 0.7%; Score 17; DB 4; Length 52297;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
Db 20184 TCTGGGACATCGAGGAC 20200

RESULT 13
US-08-705-557-1
; Sequence 1, Application US/08705557
; Patent No. 630061
; GENERAL INFORMATION:
; APPLICANT: William R. Jacobs, Jr.
; APPLICANT: Barry R. Bloom
; APPLICANT: Graham F. Hatfull
; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:

SEQUENCE CHARACTERISTICS:
LENGTH: 52297
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: phage genome sequence
DESCRIPTION: phage genome sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: not applicable.
ORIGINAL SOURCE:
ORGANISM: mycobacteriophage L5
STRAIN: not applicable
DEVELOPMENTAL STAGE: not applicable
HAPLOTYPE: not applicable
TISSUE TYPE: not applicable
CELL LINE: not applicable
ORGANELLE: not applicable
IMMEDIATE SOURCE: mycobacteriophage L5 particles
POSITION IN GENOME: entire genome
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Hatfull and Sarkis
TITLE: DNA Sequence, Structure and Gene
TITLE: Expression of Mycobacteriophage L5:
TITLE: A Phage System for Mycobacterial
TITLE: Genetics
JOURNAL: Molecular Microbiology
VOLUME: 7
PAGES: 395-405
DATE: 1993
US-09-426-436-1

Query Match 0.7%; Score 17; DB 4; Length 52297;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1924 tctgggacatcgaggac 1940
Db 20184 TCTGGGACATCGAGGAC 20200

RESULT 14
US-09-036-987A-1/c
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
```

APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
TITLE OF INVENTION: Production  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80161 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-036-987A-1

Query Match 0.7%; Score 17; DB 3; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1013 tggcaaggtcgtcgccg 1029  
|||||

Db 48698 TGGCAAGGTCTGTCGCCG 48682

RESULT 15  
US-09-370-700-1/c  
Sequence 1, Application US/09370700  
Patent No. 6274350  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H  
APPLICANT: Broughton, Mary C  
APPLICANT: Crawford, Kathryn P  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Treadway, Patti J  
APPLICANT: Turner, Jan R  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
FILE REFERENCE: 50489 DIV1  
CURRENT APPLICATION NUMBER: US/09/370,700  
CURRENT FILING DATE: 1999-08-09  
EARLIER APPLICATION NUMBER: US 09/36987  
EARLIER FILING DATE: 1998-03-09  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 80161

; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-1

Query Match 0.7%; Score 17; DB 4; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1013 tggcaaggtcgtcgccg 1029  
|||||

Db 48698 TGGCAAGGTCTGTCGCCG 48682

Search completed: December 4, 2001, 18:38:47  
Job time: 6010 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 15:08:31 ; Search time 1960.54 seconds  
(without alignments)  
12705.028 Million cell updates/sec

Title: us-09-426-072-1

Perfect score: 2318  
Sequence: 1 aattctggacacatcccc.....gcaccccgatataatt 2318

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: EST.\*

1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: 9b\_est1:\*  
11: 9b\_est2:\*  
12: 9b\_hic:\*  
13: 9b\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description
1	23	1.0	472	10	BE586656 WHE0509_A
2	21	0.9	533	13	AQ706057 HS_5550_A
3	20	0.9	420	13	AZ701728 RPCI-23-1
4	20	0.9	452	13	AQ995660 RPCI-23-3
5	20	0.9	464	11	BI322812 kx16e05.y
6	20	0.9	530	10	AI323058 mj51h08.y
7	19	0.8	138	10	AA953538 on80d06.s
8	19	0.8	151	10	AW081841 xb55h09.x
9	19	0.8	159	10	AW664184 hi05f07.x
10	19	0.8	164	11	BI063061 IL3-UT011
11	19	0.8	178	13	AZ491723 LM0325024
12	19	0.8	205	11	BF543903 UI-R-Y0-a

13	19	0.8	212	10	AW176306
c 14	19	0.8	220	10	AI203319
c 15	19	0.8	220	10	AI695918
c 16	19	0.8	223	11	BI061646
17	19	0.8	223	11	BI063217
18	19	0.8	223	11	BI063298
19	19	0.8	223	11	BI063327
c 20	19	0.8	223	11	BI063392
21	19	0.8	227	11	N56396
22	19	0.8	229	11	BI062159
c 23	19	0.8	229	11	BI062404
24	19	0.8	229	11	BI062466
25	19	0.8	229	11	BI063909
c 26	19	0.8	237	11	BI061709
c 27	19	0.8	238	10	AI197858
c 28	19	0.8	239	11	BI063783
c 29	19	0.8	247	11	BI061955
30	19	0.8	247	11	BI062259
31	19	0.8	247	11	BI062982
c 32	19	0.8	248	11	BI037052
c 33	19	0.8	255	11	BI063770
34	19	0.8	260	10	BB350572
35	19	0.8	266	13	AZ908773
c 36	19	0.8	272	10	AA652961
37	19	0.8	280	11	BI010685
c 38	19	0.8	282	10	AA927182
c 39	19	0.8	291	10	AA635619
40	19	0.8	295	11	T35546
41	19	0.8	296	10	BB347933
42	19	0.8	299	10	AI909043
c 43	19	0.8	302	10	AA809401
c 44	19	0.8	304	11	D69313
c 45	19	0.8	322	11	Z25166

#### ALIGNMENTS

#### RESULT 1

BE586656

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

location/Qualifiers

1. 472

/cultivar="Blanco"

/organism="Secale cereale"

/cultur="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

BE586656 472 bp mRNA EST 17-AUG-2000  
WHE0509\_A09\_A172R Secale cereale aluminum-stressed root tip cDNA  
library Secale cereale cDNA clone WHE0509\_A09\_A17, mRNA sequence.

BE586656

BE586656.1 GI:9839688

EST.

Secale cereale

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Secale.

1 (bases 1 to 472)

Anderson, O.D., Butler, E., Chao, S., Gustafson, J.P., Han, P.S., Hsiao,

, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and

Tong, J.C.

The structure and function of the expressed portion of the wheat

genomes - Aluminum-stressed root tip cDNA library from rye (Secale

cereale)

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: M13 reversed primer.

location/Qualifiers

1. 472

/cultivar="Blanco"

/organism="Secale cereale"

/cultivar="Blanco"

/cultivar="Blanco"

/cultivar="Blanco"

```

/db_xref="taxon:4550"
/clone="WHE0509_A09_A17"
/clone_lib="Secale cereale aluminum-stressed root tip cDNA
library"
/tissue_type="Root tip"
/dev_stage="Seedling"
/lab_host="E. coli DH12s"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Seeds
were germinated aseptically on filter paper and
transferred to a hydroponic growth system in a growth
chamber when the primary root was 1 cm in length. After a
2-day establishment period, seedlings were subjected to a
5 ppm aluminum stress prior to tissue harvest. Plants
were grown in an environmental chamber. The tissue, total
RNA, and poly(A) RNA were prepared, and a cDNA library
was made (Butler and Gustafson) at University of Missouri
, Columbia. Plasmid DNA preparations and DNA sequencing
were performed in the Od Anderson lab (all other authors
)."
BASE COUNT      94 a 164 c 120 g  94 t
ORIGIN

```

```

Query Match      1.0%; Score 23; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.62; 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0;
QY 1680 ggtgcttctacgaccatgtccc 1702
      |||||
Db 346 GGTGCTTCTACGACCATGTCCC 368

```

```

RESULT# 2
AQ706057/c
LOCUS      AQ706057      533 bp      DNA      GSS      07-JUL-1999
DEFINITION HS_5550_AI_C07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1126 Col=13 Row=E, DNA sequence.
VERSION     AQ706057.1 GI:5415483
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 533)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
            99380589
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1126 row: E column: 13
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 533.
            Location/Qualifiers
            1..533
            /organism="Homo sapiens"
            /db_xref="taxon:9606"

```

```

FEATURES
source
1..533
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="Plate=1126 Col=13 Row=E"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methyase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      183 a  81 c  74 g 181 t 14 others
ORIGIN

```

```

Query Match      0.9%; Score 21; DB 13; Length 533;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2222 ttgataacatgggatttaaaa 2242
      |||||
Db 99 TTGATACATGGGATTAAAA 79

```

```

RESULT# 3
AZ701728/c
LOCUS      AZ701728      420 bp      DNA      GSS      24-JAN-2001
DEFINITION RPCI-23-19611.TJ RPCI-23 Mus musculus genomic clone RPCI-23-19611,
            DNA sequence.
VERSION     AZ701728.1 GI:12424417
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 420)
            Zhao,S., Nieman,W., Feldblyum,T., Malek,J., Shatsman,S., Akhret
            and Fraser,C.M.
            B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
            Other GSSs: RPCI-23-19611.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
            Plate: 196 row: I column: 1
            Seq primer: SP6
            Class: BAC ends.
            Location/Qualifiers
            1..420
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-19611"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
            EcoRI; Site_2: EcoRI; female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methyase. Size
            selected DNA was cloned into the pBACE3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
            115 a  95 c 105 g 105 t
            BASE COUNT
            ORIGIN

```

```

FEATURES
source
1..420
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-19611"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
115 a  95 c 105 g 105 t
BASE COUNT
ORIGIN

```

Query Match 0.9%; Score 20; DB 13; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 gatgaactgagatgctaagc 61  
 |||||  
 Db 189 GATGAACCTGAGATGCTAAGC 170

RESULT 4  
 LOCUS AQ995660 452 bp DNA 24-FEB-2000  
 DEFINITION RPCI-23-355M11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-355M11  
 , DNA sequence.

ACCESSION AQ995660  
 VERSION AQ995660.1 GI:7070757  
 KEYWORDS GSS.  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 452)  
 REFERENCE Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akimret  
 , B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
 and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-23

TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: RPCI-23-355M11.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
 or from Resea ch Genetics ([inforesgen.com](http://inforesgen.com)). BAC end page:  
[http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 355 row: M column: 11  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES source  
 Location/Qualifiers  
 1. 452  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-355M11"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
 EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 93 a 50 c 98 g 211 t  
 ORIGIN

Query Match 0.9%; Score 20; DB 13; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 cttcatattgtttttttgt 242  
 |||||  
 Db 317 CTTCAATATTGTTTCTTTGT 336

RESULT 5  
 LOCUS BI322812/c  
 DEFINITION

BI322812 464 bp mRNA 30-JUL-2001  
 Kx16e05.y3 Parastromyoloides trichosuri IL pAMP1 v1 Chiapelli  
 McCarter Parastromyoloides trichosuri cDNA 5' similar to TR:O62515  
 O62515 PROBABLE PEROXISOMAL PHYTANOYL-COA ALPHA-HYDROXYLASE ;, mRNA  
 sequence.

ACCESSION BI322812  
 VERSION BI322812.1 GI:15001998  
 KEYWORDS EST.  
 SOURCE Parastromyoloides trichosuri.

ORGANISM Parastromyoloides trichosuri  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Panagrolaimoidea; Strongyloidea; Parastromyoloides.  
 1 (bases 1 to 464)

REFERENCE McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wyllie, T.,  
 Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., R.  
 Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe  
 , M., Allen, M., Person, B., Svaller, T., Harvey, N., Schurk, R., Kohn, S.,  
 Shin, I., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
 Wilson, R.  
 The Washington Univ. Nematode EST Project, 1999

TITLE Unpublished (1999)  
 JOURNAL Contact: McCarter JP  
 COMMENT The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@wustl.wustl.edu  
 The library was constructed by Brandi Chiapelli and Dr. James  
 McCarter (bchiapell@wustl.wustl.edu & jmcarter@wustl.wustl.edu) at  
 Washington University, St. Louis. DNA Sequencing by: Washington  
 University Genome Sequencing Center St. Louis.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 270.

FEATURES source  
 Location/Qualifiers  
 1. 464

/organism="Parastromyoloides trichosuri"  
 /db\_xref="taxon:131310"  
 /clone\_lib="Parastromyoloides trichosuri IL pAMP1 v1  
 Chiapelli McCarter"  
 /dev\_stage="Infective Larvae"  
 /lab\_host="DH10B"  
 /note="Vector: pAMP1 (Gibco); Site\_1: NotI; Site\_2: SalI;  
 The library was constructed by Brandi Chiapelli and Dr.  
 James McCarter at Washington University, St. Louis. The  
 cDNA was made by using Dynabead oligo-dt priming (Dynal).  
 PCR based library using a modified protocol from the  
 SMART PCR cDNA Synthesis Kit from Clontech. Directionally  
 cloned into the UDG sites of pAMP1. Nematodes were  
 provided by Dr. Warwick Grant of AgResearch, New Zealand  
 (warwick.grant@agresearch.co.nz)."  
 BASE COUNT 185 a 62 c 69 g 148 t  
 ORIGIN

Query Match 0.9%; Score 20; DB 11; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 ttcattatgtttttttgtc 243  
 |||||  
 Db 68 TTCATATTGTTTCTTTGTC 49

RESULT 6  
 LOCUS AI323058  
 DEFINITION mJ51h08.y1 Soares mouse embryo NBMEL3.5 14.5 Mus musculus cDNA

clone IMAGE:479679 5' similar to gb:X73960 M.musculus mRNA for TIE receptor tyrosine kinase (MOUSE);, mRNA sequence.

ACCESSION  
AI323058  
VERSION  
AI323058.1  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 530)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lemon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:290423

TITLE  
The WashU-HMMI Mouse EST Project

JOURNAL  
Unpublished (1996)  
COMMENT  
Contact: Marra M/Mouse EST Project

WashU-HMMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:290423

This read is a RESEQUENCE of a previously sequenced mouse clone  
correct orientation)  
Putative full length read  
vector to vector length is 613  
Seq primer: -40RP from Gibco  
High quality sequence stop: 325.

Location/Qualifiers

1. 530

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="Soares mouse embryo NDME13.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer (5'

TGTTACCATCTGAGTGGAGCGCGCGGAAATTTTGTGTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne

State Univ., from 2 ]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT7T3 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M. Fatima Bonaldo.

114 a 151 c 148 g 116 t 1 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Score 20; DB 10; Length 530;

100.0%; Pred. No. 28;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 ctttcctcggcgggctcag 457

Db 481 CTTTCCTCGCGGGCTCAG 462

RESULT

AA953538/c

LOCUS

138 bp mRNA

EST

07-MAY-1998

DEFINITION

IMAGE:1562987 3', mRNA sequence.

ACCESSION

AA953538

VERSION

AA953538.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 138)

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 131.

FEATURES

Location/Qualifiers

1..138

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1562987"

/lab\_host="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site.1: Not I; Site.2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI-CGAP-GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

25 a 41 c 27 g 45 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Score 19; DB 10; Length 138;

100.0%; Pred. No. 82;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 agattataaagtgtatg 2089

Db 40 AGGATTAAAGTGATGATG 22

RESULT

AW081841/c

LOCUS

151 bp mRNA

EST

DEFINITION

xb55h09.xl NCI-CGAP\_Eso2 Homo sapiens cDNA clone IMAGE:2580257 3',

mRNA sequence.

ACCESSION

AW081841

VERSION

AW081841.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 151)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Nan Hu, M.D., Ph.D., Mark Roth, M.D., Phillip

Taylor, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

on80d06.sl Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:1562987 3', mRNA sequence.

ACCESSION

AA953538

VERSION

AA953538.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 138)

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 131.

FEATURES

Location/Qualifiers

1..138

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1562987"

/lab\_host="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site.1: Not I; Site.2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI-CGAP-GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

25 a 41 c 27 g 45 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Score 19; DB 10; Length 138;

100.0%; Pred. No. 82;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 agattataaagtgtatg 2089

Db 40 AGGATTAAAGTGATGATG 22

RESULT

AW081841/c

LOCUS

151 bp mRNA

EST

DEFINITION

xb55h09.xl NCI-CGAP\_Eso2 Homo sapiens cDNA clone IMAGE:2580257 3',

mRNA sequence.

ACCESSION

AW081841

VERSION

AW081841.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 151)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Nan Hu, M.D., Ph.D., Mark Roth, M.D., Phillip

Taylor, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: polyT not found

Seq primer: -400P from Gibco

High quality sequence stop: 118.

Location/Qualifiers

## FEATURES

source

1. .151  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="NCI\_CGAP\_Eso2"  
 /tissue\_type="squamous cell carcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: esophagus; Vector: pCMV-SPORT6; Site\_1: Sali  
 ; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo  
 dt. Average insert size 1.1 kb. Life Technologies catalog  
 #: 11502-010"

BASE COUNT 24 a 41 c 25 g 61 t

## ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggattaaaagtatgatg 2089

|||||

Db 62 AGGATTAAGTATGATG 44

## RESULT 9

AW664184/c

LOCUS

AW664184 159 bp mRNA EST 06-APR-2000

h105f07.x1 NCI\_CGAP\_GUL Homo sapiens cDNA clone IMAGE:2971429 3',

mRNA sequence.

AW664184

VERSION

AW664184.1 GI:7456725

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 159)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

Possible reversed clone: polyT not found

Seq primer: -400P from Gibco.

Location/Qualifiers

1. .159

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="NCI\_CGAP\_GUL"

/tissue\_type="2 pooled high-grade transitional cell

tumors"

## FEATURES

source

/lab\_host="DH10B"

/note="Organ: genitourinary tract; Vector: pCMV-SPORT6;  
 Site\_1: Sali; Site\_2: NotI; Cloned unidirectionally.

Primer: Oligo dt. Library constructed by Life

Technologies."

BASE COUNT 29 a 44 c 34 g 52 t

## ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggattaaaagtatgatg 2089

|||||

Db 39 AGGATTAAGTATGATG 21

## RESULT 10

BI063061

LOCUS

BI063061 164 bp mRNA EST 15-JUN-2001

IL3-UT0117-280301-528-D04 UT0117 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BI063061

VERSION

BI063061.1 GI:14470588

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 164)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202863

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=IL3&t2=IL3-UT0117-280301-528-D04&t3=2001-03-28&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 164.

Location/Qualifiers

1. .164

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="UT0117"

/dev\_stage="Adult"

/note="Organ: uterus\_tumor; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 61 a 36 c 43 g 24 t

## ORIGIN

Query Match 0.8%; Score 19; DB 11; Length 164;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 aggtatataagtgatg 2089  
 Db 146 AGGATTAAGTGATG 164

## RESULT 11

AZ491723 178 bp DNA GSS 05-OCT-2000  
 LOCUS LM0325024F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC1M0325024 F, DNA sequence.

ACCESSION AZ491723  
 VERSION AZ491723.1 GI:10663701  
 KEYWORDS GSS.  
 SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 178)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Lingey, A., von Niederhausen, A.  
 and Wright, D., Weiss, R.

## TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

## JOURNAL

## COMMENT

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0325 row: 0 column: 24

Seq primer: CGTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 178.

## FEATURES

source

Location/Qualifiers

1..178

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0325024"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/notes="Vector: FWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gbAF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

48 a 44 c 40 g 46 t

## BASE COUNT

## ORIGIN

Query Match

0.8%; Score 19; DB 13; Length 178;

Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1419 gattctctcttcacct 1437

Db 77 GATTCTCTCTTCACCT 95

## RESULT 12

BF543903

LOCUS BF543903

DEFINITION BF543903 205 bp mRNA EST 11-DEC-2000  
 UI-R-Y0-abx-g-10-0-UI-r1 UI-R-Y0 rattus norvegicus cDNA clone

ACCESSION BF543903

VERSION BF543903.1 GI:11635010

KEYWORDS EST.

SOURCE Norway rat.

## ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 205)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

cDNA library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LLNL (info@image.llnl.gov). IMAGE ID= 1791289

Seq primer: M13 Forward.

## FEATURES

source

Location/Qualifiers

1..205

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-Y0-abx-g-10-0-UI"

/clone\_lib="UI-R-Y0"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-Y0

library is a subtracted library derived from an

individually-tagged normalized whole-eye (minus the lens)

library. The driver for the subtraction consisted of a

pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,

UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of

3-5 nucleotides present between the Not I site and the

oligo-dT track which allows identification of the library

of origin of a clone within the mixture. The subtracted

library (UI-R-Y0) was constructed as follows: PCR

amplified cDNA inserts from previous library clones from

which 3' ESTs had been derived were used as a driver in a

hybridization with the normalized whole-eye library in

the form of single-stranded circles. The remaining

single-stranded circles (subtracted library) was purified

by hydroxyapatite column chromatography, converted to

double-stranded circles and electroporated into DH10B

bacteria (Life Technologies) to generate the UI-R-Y0

library. This procedure has been previously described

(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,

1996)"

52 a 37 c 52 g 64 t

## BASE COUNT

## ORIGIN

Query Match	0.88;	Score 19;	DB 11;	Length 205;
Best Local Similarity	100.0%;	Pred. No. 86;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1169	caacagctggtctctg9t 1187			
DB 78	CAACAGCTGGTTCGCTGT 96			
RESULT 13				
AW176306				
LOCUS	212 bp	mRNA	EST	16-NOV-1999
DEFINITION	QVO-BF0229-310899-011-d10 BF0229 Homo sapiens cDNA, mRNA sequence.			
ACCESSION	AW176306			
VERSION	AW176306.1	GI:6442343		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 212)			
AUTHORS	HCGP <a href="http://www.ludwig.org.br/ORESTES">http://www.ludwig.org.br/ORESTES</a> .			
TITLE	The FAPESP/LICR Human Cancer Genome Project			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: <a href="mailto:asimpson@ludwig.org.br">asimpson@ludwig.org.br</a> This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QVO&amp;t3-QVO-BT0229-310899-011-d10&amp;t3-1999-08-31&amp;t4-1">http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QVO&amp;t3-QVO-BT0229-310899-011-d10&amp;t3-1999-08-31&amp;t4-1</a> ) Seq primer: puc 18 forward High quality sequence start: 31 High quality sequence stop: 212.			
FEATURES				
source	1. .212			
	Location/Qualifiers			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_lib="BT0229"			
	/dev_stage="Adult"			
	/note="organ: breast; Vector: pUC18; Site:1; SmaI; Site:2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 136 /116 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
BASE COUNT	62 a	64 c	48 g	38 t
ORIGIN				
Query Match	0.88;	Score 19;	DB 10;	Length 212;
Best Local Similarity	100.0%;	Pred. No. 87;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 614	gatacaagcctggctccaa 632			
DB 11	GATACAAAGCCTGGCTCAA 29			
RESULT 14				
AI203319/c				
LOCUS	220 bp	mRNA	EST	14-OCT-1998
DEFINITION	qr29g06.x1 NC1_CGAP_G66 Homo sapiens cDNA clone IMAGE:1942330 3', mRNA sequence.			
ACCESSION	AI203319			
VERSION	AI203319.1	GI:3755925		

KEYWORDS	EST.	human.
SOURCE	human sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 220)	
AUTHORS	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/nciccap">http://www.ncbi.nlm.nih.gov/nciccap</a> .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps@remail.nih.gov">cgaps@remail.nih.gov</a> Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www.bio.llnl.gov/bbrp/image/image.html">www.bio.llnl.gov/bbrp/image/image.html</a> Seq primer: -40up from Gibco.	
FEATURES	Location/Qualifiers	
source	1..220	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:1942330"	
	/clone_lib="NCI-CCAP-GC6"	
	/tissue_type="pooled germ cell tumors"	
	/lab_host="DH108"	
	/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CCAP-GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	35 a 62 c 45 g 78 t	
ORIGIN		
Query Match	0.8%; Score 19; DB 10; Length 220;	
Best Local Similarity	100.0%; Pred. No. 87;	
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	2071 agattataaaagtcatgatg 2089	
Db	42 AGCAITTAAGTCATGATG 24	
RESULT	15	
AI695918/c	220 bp mRNA	EST
DEFINITION	ts87e09.x1 NCI-CCAP-GC6 Homo sapiens cDNA clone IMAGE:2238280 3', mRNA sequence.	03-JUN-1999
ACCESSION	AI695918	
VERSION	AI695918.1	
KEYWORDS	GI:4983818	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Human sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 220)	
JOURNAL	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/nciccap">http://www.ncbi.nlm.nih.gov/nciccap</a> .	
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps@remail.nih.gov">cgaps@remail.nih.gov</a> Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.	

KEYWORDS	EST.	human.
SOURCE	human sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 220)	
AUTHORS	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/nciccap">http://www.ncbi.nlm.nih.gov/nciccap</a> .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps@remail.nih.gov">cgaps@remail.nih.gov</a> Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www.bio.llnl.gov/bbrp/image/image.html">www.bio.llnl.gov/bbrp/image/image.html</a> Seq primer: -40up from Gibco.	
FEATURES	Location/Qualifiers	
source	1..220	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:1942330"	
	/clone_lib="NCI-CCAP-GC6"	
	/tissue_type="pooled germ cell tumors"	
	/lab_host="DH108"	
	/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CCAP-GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	35 a 62 c 45 g 78 t	
ORIGIN		
Query Match	0.8%; Score 19; DB 10; Length 220;	
Best Local Similarity	100.0%; Pred. No. 87;	
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	2071 agattataaaagtgtatgatg 2089	
Db	42 AGCAITTAAGTCATGTCATG 24	
RESULT	15	
AI695918/c	220 bp mRNA	EST
DEFINITION	ts87e09.x1 NCI-CCAP-GC6 Homo sapiens cDNA clone IMAGE:2238280 3', mRNA sequence.	03-JUN-1999
ACCESSION	AI695918	
VERSION	AI695918.1	GI:4983818
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	human sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 220)	
TITLE	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/nciccap">http://www.ncbi.nlm.nih.gov/nciccap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps@remail.nih.gov">cgaps@remail.nih.gov</a> Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.	

cdna Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cdna Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -400P from Gibco.

Location/Qualifiers  
1. .220  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2238280"  
/clone\_lib="NCI-CGAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 1257096-1258631, 1469064-1470983, and 1475592-1476743).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 35 a 62 c 45 g 78 t  
ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 220;  
Best Local Similarity 100.0%; Pred.No. 87; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 aggattaaaaagtgtgatg 2089  
|||||  
Db 42 AGGATTAAAGTGTGTG 24

Search completed: December 4, 2001, 17:35:19  
Job time: 8808 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 17:03:47 ; Search time 208.51 seconds  
(without alignments)  
9530.861 Million cell updates/sec

Title: us-09-426-072-1  
Sequence: 1 aattctgacacaaatcccc.....gcaccccgatacattatt 2318

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	ID	Description
1	2318	100.0	2318	22 AAA54535
2	21	0.9	30	22 AAA54537
3	20	0.9	3748	22 AAA54226
4	19	0.8	304	22 AAH54226
5	19	0.8	1108	20 AAX60581
6	19	0.8	1263	21 AAA27058
7	19	0.8	1281	21 AAA27059
8	19	0.8	3778	22 AAH16638
9	19	0.8	119950	20 AAX90201
10	18	0.8	27	22 AAA54536
11	18	0.8	40	21 AAC61921

12	18	0.8	1070	22 AAH33057
c 13	18	0.8	1224	21 AAC61895
14	18	0.8	1307	21 AA229801
15	18	0.8	1394	22 AA162830
c 16	18	0.8	1549	21 AAC69521
17	18	0.8	1766	21 AAA88846
18	18	0.8	1766	21 AA245385
19	18	0.8	1766	22 AA501105
c 20	18	0.8	2761	22 AA162712
21	18	0.8	4450	22 AAC61960
c 22	18	0.8	6832	22 AAH21139
c 23	18	0.8	8358	22 AAH21140
c 24	18	0.8	8358	20 AA232028
25	18	0.8	38186	22 AAC90085
c 26	17	0.7	160	21 AAC18914
27	17	0.7	172	20 AA286664
c 28	17	0.7	211	16 AA20160
c 29	17	0.7	252	22 AAH83697
c 30	17	0.7	253	22 AAH83030
c 31	17	0.7	300	21 AAA00141
c 32	17	0.7	345	22 AAH82747
33	17	0.7	373	19 AA286605
34	17	0.7	375	21 AAA64651
c 35	17	0.7	404	22 AAH82773
36	17	0.7	404	22 AA2767364
c 37	17	0.7	412	22 AAH83539
c 38	17	0.7	429	22 AA2766595
c 39	17	0.7	470	14 AAQ39686
c 40	17	0.7	470	14 AAQ59098
41	17	0.7	482	21 AAC75557
42	17	0.7	500	21 AAC78019
43	17	0.7	500	22 AAH33259
44	17	0.7	509	20 AA241959
45	17	0.7	513	22 AAH67214

#### ALIGNMENTS

RESULT 1  
AAA54535  
ID AAA54535 standard; DNA; 2318 BP.  
XX AC AAA54535;  
XX 11-APR-2001 (first entry)  
XX Phospholipase B coding sequence of Aspergillus oryzae.  
DE Phospholipase B; hydrolysis; hydrolyse; lysophospholipid;  
KW phospholipid; lecithin; lysolethicin; degumming; starch;  
KW wheat starch; oil; dough; baking; strength; elasticity;  
KW stability; extensibility; ds.  
XX Aspergillus oryzae.  
OS Aspergillus oryzae.  
XX Key Location/Qualifiers  
FT exon 510..696  
FT exon /\*tag= a  
FT FT /\*label= Exon 1  
FT Intron 697..788  
FT FT /\*tag= b  
FT FT /\*label= Intron 1  
FT exon 789..1196  
FT FT /\*tag= c  
FT FT /\*label= Exon 2  
FT Intron 1197..1247  
FT FT /\*tag= d  
FT FT /\*label= Intron 2  
FT exon 1248..2048  
FT FT /\*tag= e  
FT FT /\*label= Exon 3  
XX

US6146869-A.  
PD 14-NOV-2000.  
XX 21-OCT-1999; 99US-0426072.  
XX 21-OCT-1999; 99US-0426072.  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
XX Harris P, Brown KM;  
XX WPI; 2001-049057/06.  
DR P-PSDB; AAB04146.  
XX Novel polypeptide having phospholipase B activity useful for  
PT hydrolysing fatty acyl groups of phospholipid, preparing dough or baked  
PT product and reducing phospholipid content in an edible oil  
XX Claim 1; Fig 1a-b; 24pp; English.  
CC Phospholipase B is useful for hydrolysing the fatty acyl group(s) of  
CC phospholipids or lysophospholipids, such as lecithin or lysolecithin,  
CC degumming an aqueous carbohydrate solution or slurry to improve its  
CC filterability, particularly, a starch hydrolysate or wheat starch  
CC hydrolysate which is difficult to filter and to reduce the  
CC phospholipid content in an edible oil by treating the oil with  
CC phospholipase B to hydrolyse a major portion of the phospholipid  
CC and separating an aqueous phase containing the hydrolysed  
CC phospholipid from the oil. Phospholipase B is also useful for  
CC preparing a dough or a baked product having improved property such  
CC as increased strength, elasticity, stability and extensibility.  
XX Sequence 2318 BP; 554 A; 655 C; 534 G; 575 T; 0 other;  
SQ  
Query Match 100.0%; Score 2318; DB 22; Length 2318;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 aattctgtgacacaaatcccttgagagtgatagtagggggatgaactgagatgctaaag 60  
DB 1 aattctgtgacacaaatcccttgagagtgatagtagggggatgaactgagatgctaaag 60  
QY 61 ctacagttcttcgcttgccgttaggctgcccattgttgggttaggtctctctccacc 120  
DB 61 ctacagttcttcgcttgccgttaggctgcccattgttgggttaggtctctctccacc 120  
QY 121 ctadtagtgatgcaattgctgttccaaatcccatcccatctcgatcagctctattat 180  
DB 121 ctadtagtgatgcaattgctgttccaaatcccatcccatctcgatcagctctattat 180  
QY 181 gaccaagcgtgtcaaatgagagcccaactcatgtggcctccacttcatattgtttcttt 240  
DB 181 gaccaagcgtgtcaaatgagagcccaactcatgtggcctccacttcatattgtttcttt 240  
QY 241 gtcgataagactgtcatccctgtccggtggttccacgagaaacattgttaagatgg 300  
DB 241 gtcgataagactgtcatccctgtccggtggttccacgagaaacattgttaagatgg 300  
QY 301 atagcgtataggctgcccctgttcggtatgaagccgaagccctcgtcgtcgatggatat 360  
DB 301 atagcgtataggctgcccctgttcggtatgaagccgaagccctcgtcgtcgatggatat 360  
QY 361 gcggttcaaccgacgacagcagcagctggtatccgaaatggagcgaaggttagctctc 420  
DB 361 gcggttcaaccgacgacagcagcagctggtatccgaaatggagcgaaggttagctctc 420  
QY 421 cagtggtgctataaagccttccctggccggggtcagactcccttagcagcagcattggt 480  
DB 421 cagtggtgctataaagccttccctggccggggtcagactcccttagcagcagcattggt 480  
QY 481 caagctcacgacctcaatcccgagtagaaacatgaagctccaccgctctgttactggtctc 540

DB 481 caagctcacgacctcaatcccgagtagaaacatgaagctccaccgctgttactggtctc 540  
QY 541 ggcctcttggcctctctcggtcttctagcctgttcacgtccgagtagtacgagcgtgcga 600  
DB 541 ggcctcttggcctctctcggtcttctagcctgttcacgtccgagtagtacgagcgtgcga 600  
QY 601 gaagcccttctcgatatacaagcctggctcaagagtagtccattgagaacttgaagtagcaag 660  
DB 601 gaagcccttctcgatatacaagcctggctccaaaggagtcattgagaacttgaagtagcaag 660  
QY 661 gtcgagaacattgtctggcttattctcgagaaacaggtttgtgtgccccttacgattatcat 720  
DB 661 gtcgagaacattgtctggcttattctcgagaaacaggtttgtgtgccccttacgattatcat 720  
QY 721 atgtggaataataaattctcaattcagctgtgtgtatgtaagacgagcactaaacat 780  
DB 721 atgtggaataataaattctcaattcagctgtgtgtatgtaagacgagcactaaacat 780  
QY 781 ggtccagatccttcgataacattctggaggtgctgcgcgccaagagtaggacaaccgga 840  
DB 781 ggtccagatccttcgataacattctggaggtgctgcgcgccaagagtaggacaaccgga 840  
QY 841 tcaacaacgcccgttctcgaactacaagaatgcgagacccatctcgcgggaagtact 900  
DB 841 tcaacaacgcccgttctcgaactacaagaatgcgagacccatctcgcgggaagtact 900  
QY 901 gtactcagggccaaggactatgattccgtgttcaacgactccagaccactcgcgtgactgta 960  
DB 901 gtactcagggccaaggactatgattccgtgttcaacgactccagaccactcgcgtgactgta 960  
QY 961 ataaacttgagttctcgaagacttacacccaacaatggtgcgattgagtggaag 1020  
DB 961 ataaacttgagttctcgaagacttacacccaacaatggtgcgattgagtggaag 1020  
QY 1021 tggctgcgcgacagctgtgcttctcgaacgacagcttaacgactaccacaactgccc 1080  
DB 1021 tggctgcgcgacagctgtgcttctcgaacgacagcttaacgactaccacaactgccc 1080  
QY 1081 cagaagagcgcagcaagcaagtgatggatactatacagagagagaggttctactacgtcg 1140  
DB 1081 cagaagagcgcagcaagcaagtgatggatactatacagagagagaggttctactacgtcg 1140  
QY 1141 tggacctgtgagtgagttcaacttcaacagctgttctcgtgtgttctcgtggtgta 1200  
DB 1141 tggacctgtgagtgagttcaacttcaacagctgttctcgtgtgttctcgtggtgta 1200  
QY 1201 gtgcgataaattctgtatttttaattttgactaacgcttcaaaagcctaccacaacca 1260  
DB 1201 gtgcgataaattctgtatttttaattttgactaacgcttcaaaagcctaccacaacca 1260  
QY 1261 accgtgtgtgctctcgaagacacgctgctggcagtggaagatgacgagacttcc 1320  
DB 1261 accgtgtgtgctctcgaagacacgctgctggcagtggaagatgacgagacttcc 1320  
QY 1321 tgaactatgttctctagcaagtcattcttcgagggcgcgcaacgagaagggcgttccct 1380  
DB 1321 tgaactatgttctctagcaagtcattcttcgagggcgcgcaacgagaagggcgttccct 1380  
QY 1381 ggcctcaactcagctgacacacgagagatcgaacgggattctcttcttccactacg 1440  
DB 1381 ggcctcaactcagctgacacacgagagatcgaacgggattctcttcttccactacg 1440  
QY 1441 tcaaccagacctcccggttccaaacgtgtgtcccggttgaaaacttcttccaaagcgttacc 1500  
DB 1441 tcaaccagacctcccggttccaaacgtgtgtcccggttgaaaacttcttccaaagcgttacc 1500  
QY 1501 tccggtctctccctaaattctcttaccattacccctctcgtcgtcgccacacacccaact 1560  
DB 1501 tccggtctctccctaaattctcttaccattacccctctcgtcgtcgccacacacccaact 1560  
QY 1561 ccatgcaccccccacgggtaacgctctcactcaggttgaggttctcgtcaagcagatctatg 1620

Db	1561	ccatgcacccacccggttaacgctctctctacggttgaggcttctcgtcaagcagatctctatgctg	1620
QY	1621	ccatttcgcaggccctcagtggtggacgaacccctgctcttattactacgacgagaccg	1680
Db	1621	ccatttcgcaggccctcagtggtggacgaacccctgctcttattactacgacgagaccg	1680
QY	1681	gtggctttacgacacatgttcctccctctcgcgttcgcgcggacacactgcacctaca	1740
Db	1681	gtggctttacgacacatgttcctccctctcgcgttcgcgcggacacactgcacctaca	1740
QY	1741	ctgagactgcgaagaacggctcgaagaatacactcttcaattcgacccgtctgggtgcgcga	1800
Db	1741	ctgagactgcgaagaacggctcgaagaatacactcttcaattcgacccgtctgggtgcgcga	1800
QY	1801	tgcgcacctgggttatctccctctacagtgaagaaggatatactgcgcagtcagcgaacgg	1860
Db	1801	tgcgcacctgggttatctccctctacagtgaagaaggatatactgcgcagtcagcgaacgg	1860
QY	1861	atccgctcaacggcgaagcccgctccctacagtgctctacotccgctctcaagaactctcgagat	1920
Db	1861	atccgctcaacggcgaagcccgctccctacagtgctctacotccgctctcaagaactctcgagat	1920
QY	1921	atctctgggacatcgcaggaccttcaacctcgtgtctgcaccaactctccattcttcgatacc	1980
Db	1921	atctctgggacatcgcaggaccttcaacctcgtgtctgcaccaactctccattcttcgatacc	1980
QY	1981	tgatctggacacactcttcgttgaggatgctctctatgtctctcaagaactccccatacctttt	2040
Db	1981	tgatctggacacactcttcgttgaggatgctctctatgtctctcaagaactccccatacctttt	2040
QY	2041	cgggtataagctctcagtcctcgtgcagtgagacagattataaagtgcgatgcacgttctgcac	2100
Db	2041	cgggtataagctctcagtcctcgtgcagtgagacagattataaagtgcgatgcacgttctgcac	2100
QY	2101	ttcagtgacaacttaccatgttatagacatgttttgcgtctatagctaccgcctagagcgag	2160
Db	2101	ttcagtgacaacttaccatgttatagacatgttttgcgtctatagctaccgcctagagcgag	2160
QY	2161	cgcgcgtatgataagaataaagctgggttattctctattgttatattcaatgaagaagac	2220
Db	2161	cgcgcgtatgataagaataaagctgggttattctctattgttatattcaatgaagaagac	2220
QY	2221	tttgataaacatgggattataaagaagaatgttttctgcacaactcaactcagcgaacagg	2280
Db	2221	tttgataaacatgggattataaagaagaatgttttctgcacaactcaactcagcgaacagg	2280
QY	2281	cgttaactttcacgtgcagtcgaccccggtacattaatt	2318
Db	2281	cgttaactttcacgtgcagtcgaccccggtacattaatt	2318

## RESULT 2

RESULT 2  
AAA54537/C

AAA54537/C  
ID AAA54537 standard: DNA: 30 BP.

XX  
XX  
XXXXXX

100

AC AAA54537;

XX

DT 11-APR-20

XX 25

DE Primer fo  
vvXX  
KW  
Phosphor i

KW Phosphorl

KW phosphorus  
KW wheat straw

KW wheel stability

[illegible]

OS Aspergill

[illegible]

PS Claim 8; Page 1180-1182; 2188pp; English.  
 XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II) given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the amplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX Sequence 3748 BP; 1287 A; 569 C; 650 G; 1242 T; 0 other;

Query Match 0.9%; Score 20; DB 22; Length 3748;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 220 ccacttcattgttttctt 239  
 Db 2250 CCACCTCATATGTTTCTT 2231

## RESULT 4

AAZ77482  
 ID AAZ77482 standard; cDNA; 304 BP.  
 AC AAZ77482;  
 DT 10-APR-2000 (first entry)  
 DE Human ovarian tumor cDNA library derived EST fragment 33.  
 KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;  
 KW gene therapy; treatment; ss.  
 XX Homo sapiens.  
 OS DE19817557-Al.  
 PN 21-OCT-1999.  
 PD 09-APR-1998; 98DE-1017557.  
 XX 09-APR-1998; 98DE-1017557.  
 PR (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 PI WPI: 1999-591920/51.  
 DR P-PSDB; AA76578, AA76579, AA76580.  
 XX New nucleic acid sequences expressed in ovarian, and some other, cancer  
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and  
 PT identification of therapeutic agents -  
 XX Claim 3; Page 164; 310pp; German.  
 PS This invention describes novel nucleic acid (cDNA) sequences (A) which  
 CC have anticancer activity and are highly expressed in ovarian tumor

CC tissue (and some also in testis and breast cancer tissue). The products  
 CC of the invention can be used for gene therapy. (A) are used (i) for  
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete  
 CC genes. (B) are used (i) to identify agents suitable for treatment of  
 CC ovarian cancer; (ii) directly for treating this form of cancer  
 CC (including expression from gene therapy vectors) and (iii) for generation  
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
 CC sequence tags) from a particular tissue type before comparison of the  
 CC expression patterns. This allows a significantly longer fragment of the  
 CC gene to be revealed, so should reduce the number of failures associated  
 CC with the fact that ESTs from different libraries may represent different  
 CC parts of the same unknown gene, distorting the estimated frequency of  
 CC occurrence in a particular tissue. AA77450-777572 represent the human  
 CC ovarian tumor cDNA library derived EST fragments described in the method  
 CC of the invention and encode the protein fragments represented in  
 CC AA76505-76638.

XX Sequence 304 BP; 88 A; 66 C; 94 G; 56 T; 0 other;

Query Match 0.8%; Score 19; DB 20; Length 304;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 807 ggaggcgtgcgcgccaag 825

Db 207 ggaggcgtgcgcgccaag 225

## RESULT 5

AAZ60581  
 ID AAZ60581 standard; DNA; 1108 BP.

XX AAZ60581;  
 AC AAZ60581;  
 DT 27-JUL-1999 (first entry)

DE Human secreted protein encoding DNA (clone bhl57\_7).  
 KW Secreted protein; human; tissue marker; genetic disease; gene therapy;  
 KW veterinary medicine; cell proliferation; immunostimulant; infection;  
 KW immunosuppressant; autoimmune disease; organ rejection; tumour; anaemia;  
 KW haematopoiesis; wound healing; fertility control; chemotaxis; analgesic;  
 KW thrombolytic; haemophilia; infarction; antimicrobial agent; cancer; ss.

XX Homo sapiens.  
 OS WO9244469-Al.  
 PN 20-MAY-1999.  
 PD 06-NOV-1998; 98WO-US23829.  
 XX 04-NOV-1998; 98US-0185936.  
 PR 07-NOV-1997; 97US-0965789.  
 XX (GEM) GENETICS INST INC.  
 PA Acostino MJ, Evans C, Jacobs K, Lavallie ER, McCoy JM;  
 PI Merberg D, Racie LA, Treacy M;  
 XX WPI: 1999-327362/27.  
 DR P-PSDB; AA716781.  
 XX Nucleic acid encoding secreted human proteins  
 PT Claim 19; Page 93; 107pp; English.

CC The invention provides polynucleotides (AAZ60579-X60687) encoding  
 CC specific secreted human proteins (AA716779-Y16787). The nucleic acid  
 CC sequences are deposited under the accession number ATCC 98580. The  
 CC polynucleotides are used as tissue markers, chromosomal tags, for  
 CC diagnosis of genetic diseases, to generate anti-protein or anti-DNA

CC antibodies, also as nutritional sources and supplements and in gene  
 CC therapy. The secreted proteins are useful therapeutically, in human or  
 CC veterinary medicine, e.g. for modulating cell proliferation or  
 CC differentiation, as immunostimulants or immunosuppressants (for treating  
 CC infections, autoimmune disease, organ rejection, or to induce tumour  
 CC immunity), as regulators of haematopoiesis (e.g. for treating anemia or  
 CC in conjunction with tumour therapy), to stimulate growth of tissue for  
 CC wound healing, as fertility control agents, for regulating chemotaxis or  
 CC chemokines (e.g. for directing cells to tumours or sites of infection), as  
 CC haemostatic and thrombolytic agents (e.g. in treatment of haemophilia or  
 CC infarctions), as antimicrobial agents, for modifying biorhythms,  
 CC appetite, or metabolism, as analgesics and many other uses. The proteins  
 CC are also used to raise antibodies, used as diagnostic immunoassay  
 CC reagents also (when neutralizing) for treating e.g. cancer.  
 CC  
 SQ Sequence 1108 BP; 441 A; 191 C; 299 G; 177 T; 0 other;

Query Match 0.8%; Score 19; DB 20; Length 1108;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 aggattaaaagtgatgatg 2089  
 |||||  
 DB 907 aggtataaaagtgatgatg 925

RESULT 6  
 AAAA27058  
 ID AAAA27058 standard; DNA; 1263 BP.  
 AC AAAA27058;  
 XX  
 XX 22-AUG-2000 (first entry)  
 XX Human 5T4 tumour-associated antigen gene.  
 XX  
 XX Human: TAA; tumour-associated antigen; anti-tumour; cytostatic;  
 KW immunostimulant; vaccine; carcinoma; colorectal cancer;  
 KW gastric cancer; ds.  
 XX Homo sapiens.  
 XX WO200029428-A2.  
 XX 25-MAY-2000.  
 XX 18-NOV-1999; 99WO-GB03859.  
 XX 18-NOV-1998; 98GB-0025303.  
 PR 27-JAN-1999; 99GB-0001739.  
 PR 30-JUL-1999; 99GB-0017995.  
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.  
 XX Carroll MW, Myers KA;  
 XX WPI; 2000-387735/33.  
 XX Tumor associated antigen, 5T4 capable of eliciting cytotoxic  
 PT T-lymphocyte response useful in vaccinating against and in treating  
 PT tumors.  
 XX Example 2; Page 78; 79pp; English.  
 XX The present sequence encodes the human 5T4 tumour-associated  
 CC antigen (TAA). The TAA 5T4 is a glycoprotein which is widely  
 CC expressed in carcinomas but has a highly restricted expression pattern  
 CC in normal adult tissues. It appears to be strongly correlated to  
 CC metastasis in colorectal and gastric cancer. 5T4 antigen may therefore  
 CC be useful in tumour diagnosis, targeting and immunotherapy.  
 CC Mice in which tumours had been induced were inoculated with a virus  
 CC expression vector containing the present sequence. The 5T4 antigen  
 PS was shown to be effective at eliciting an immunotherapeutic anti-tumour  
 CC response. Both the nucleic acid encoding the antigen and the antigen  
 CC itself can be used to elicit an immune response, preferably CTL or an  
 CC antibody response in a subject. The present sequence appears in  
 CC GenBank at accession number AF012160.  
 XX Sequence 1281 BP; 246 A; 410 C; 354 G; 271 T; 0 other;

CC was shown to be effective at eliciting an immunotherapeutic anti-tumour  
 CC response. Both the nucleic acid encoding the antigen and the antigen  
 CC itself can be used to elicit an immune response, preferably CTL or an  
 CC antibody response in a subject.  
 XX  
 SQ Sequence 1263 BP; 230 A; 428 C; 349 G; 256 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 1263;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttctgtggacacaatccct 21  
 |||||  
 DB 868 ttctgtggacacaatccct 886

RESULT 7  
 AAAA27059  
 ID AAAA27059 standard; DNA; 1281 BP.  
 AC AAAA27059;  
 XX  
 XX 22-AUG-2000 (first entry)  
 XX Mouse 5T4 tumour-associated antigen gene.  
 XX  
 XX Mouse: TAA; tumour-associated antigen; anti-tumour; cytostatic;  
 KW immunostimulant; vaccine; carcinoma; colorectal cancer;  
 KW gastric cancer; ds.  
 XX Mus musculus.  
 XX WO200029428-A2.  
 XX 25-MAY-2000.  
 XX 18-NOV-1999; 99WO-GB03859.  
 XX 18-NOV-1998; 98GB-0025303.  
 PR 27-JAN-1999; 99GB-0001739.  
 PR 30-JUL-1999; 99GB-0017995.  
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.  
 XX Carroll MW, Myers KA;  
 XX WPI; 2000-387735/33.  
 XX Tumor associated antigen, 5T4 capable of eliciting cytotoxic  
 PT T-lymphocyte response useful in vaccinating against and in treating  
 PT tumors.  
 XX Example 2; Page 78; 79pp; English.  
 XX The present sequence encodes the mouse 5T4 tumour-associated  
 CC antigen (TAA). The TAA 5T4 is a glycoprotein which is widely  
 CC expressed in carcinomas but has a highly restricted expression pattern  
 CC in normal adult tissues. It appears to be strongly correlated to  
 CC metastasis in colorectal and gastric cancer. 5T4 antigen may therefore  
 CC be useful in tumour diagnosis, targeting and immunotherapy.  
 CC Mice in which tumours had been induced were inoculated with a virus  
 CC expression vector containing the present sequence. The 5T4 antigen  
 CC was shown to be effective at eliciting an immunotherapeutic anti-tumour  
 CC response. Both the nucleic acid encoding the antigen and the antigen  
 CC itself can be used to elicit an immune response, preferably CTL or an  
 CC antibody response in a subject. The present sequence appears in  
 CC GenBank at accession number AF012160.  
 XX Sequence 1281 BP; 246 A; 410 C; 354 G; 271 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 1281;

Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttcttggaacaatacct 21  
J|||||J|||||J|||||J  
Db 886 ttcttggaacaatacct 904

RESULT 8  
AAH16638/c  
ID AAH16638 standard; CDNA; 3778 BP.

XX AC AAH16638;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:15756.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EPI074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PS Claim 8; SEQ ID 15756; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602  
XX CC full-length cDNAs defined in the specification. Where a primer set  
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX CC to the complementary strand of a polynucleotide which comprises one of  
XX CC the 5602 nucleotide sequences defined in the specification, where the  
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX CC of an oligonucleotide comprising a sequence complementary to the  
XX CC complementary strand of a polynucleotide which comprises a 5'-end  
XX CC sequence and an oligonucleotide comprising a sequence complementary to a  
XX CC polynucleotide which comprises a 3'-end sequence, where the  
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of  
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in  
XX CC the specification. The primer sets can be used in antisense therapy and  
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX CC particularly full-length cDNAs. The primers are also useful for the  
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by  
XX CC the full-length cDNAs. The primers allow obtaining of the full-length  
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX CC represent oligonucleotides, all of which are used in the exemplification  
XX CC of the present invention.

XX Sequence 3778 BP; 773 A; 982 C; 1104 G; 919 T; 0 other;

Query Match 0.8%; Score 19; DB 22; Length 3778;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1447 agacctccggtccaact 1465

Db 1424 AGACCTCCCGGTCAACGT 1406

RESULT 9  
AAH90201/c  
ID AAX90201 standard; DNA; 119950 BP.

XX AC AAX90201;

XX DT 23-SEP-1999 (first entry)

XX DE Human yesl gene.

XX KW Human; yesl; diagnosis; neuropsychiatric disorder; BAD; schizophrenia;  
XX KW bipolar affective disorder; attention deficit disorder;  
XX KW schizoaffective disorder; unipolar affective disorder;  
XX KW Huntington's disease; Parkinson's disease; manic-depression; ds.

XX OS Homo sapiens.

XX PN WO9935290-A1.

XX PD 15-JUL-1999.

XX PF 07-JAN-1999; 99WO-US00297.

XX PR 08-JAN-1998; 98US-0003944.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Chen H, Freilmer NB;

XX DR WPI; 1999-444203/37.

XX DR P-PSDB; AAY24421.

XX PT Detection of a genetic mutation in the yesl gene, useful for

XX PT diagnosis of a yesl mediated neuropsychiatric disorder in a human

XX PS Claim 1; Fig 2; 110pp; English.

XX CC The present invention describes a method for detecting a genetic  
XX CC mutation in the yesl gene for the diagnosis of a yesl mediated  
XX CC neuropsychiatric disorder in a human. The method comprises detecting the  
XX CC presence or absence of a genetic mutation in the yesl gene of the  
XX CC subject, where the genetic mutation is a substitution, insertion or a  
XX CC deletion and results in the production of a yesl protein having an amino  
XX CC acid sequence other than the wild-type yesl amino acid sequence and the  
XX CC presence of the genetic mutation identifies a subject that has or is at  
XX CC risk for developing a yesl mediated neuropsychiatric disorder. Compounds  
XX CC that bind to the yesl protein, alter the amount of the protein, or alter  
XX CC the activity of the yesl gene product, are useful for treating a yesl  
XX CC mediated neuropsychiatric disorder. The disorders include Huntington's  
XX CC disease, Parkinson's disease, and especially bipolar affective disorder  
XX CC (BAD) also known as bipolar mood disorder (BP) or manic-depressive  
XX CC illness. The method distinguishes neuropsychiatric disorders from  
XX CC neurological disorders, which enables more accurate evaluation and  
XX CC prescription of medical treatment. The present sequence represents the  
XX CC human yesl cDNA sequence.

XX Sequence 119950 BP; 34471 A; 23730 C; 24660 G; 37033 T; 56 other;

Query Match 0.8%; Score 19; DB 20; Length 119950;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1208 aaatctgtatattttaat 1226
Db 3945 AAATCTGTATATTTTAAAT 3927

RESULT 10
AAAS4536
ID AAAS4536 standard; DNA; 27 BP.
XX
AC AAAS4536;
XX
XX 11-APR-2001 (first entry)
XX
XX Primer for amplifying phospholipase B coding sequence.
XX
XX Phospholipase B; hydrolysis; hydrolyse; lysophospholipid;
XX phospholipid; lecithin; lysolecithin; degumming; starch;
XX wheat starch; oil; dough; baking; strength; elasticity;
XX stability; extensibility; primer; ss.
XX
XX Aspergillus oryzae.
XX
XX US6146869-A.
XX
XX 14-NOV-2000.
XX
XX 21-OCT-1999; 99US-0426072.
XX
XX 21-OCT-1999; 99US-0426072.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX
XX Harris P, Brown KM;
XX
XX WPI; 2001-049057/06.
XX
XX Novel polypeptide having phospholipase B activity useful for
XX hydrolysing fatty acyl groups of phospholipid, preparing dough or baked
XX product and reducing phospholipid content in an edible oil
XX
XX Example 3; Column 31; 24pp; English.
XX
XX Phospholipase B is useful for hydrolysing the fatty acyl group(s) of
XX phospholipids or lysophospholipids, such as lecithin or lysolecithin,
XX degumming an aqueous carbohydrate solution or slurry to improve its
XX filterability, particularly, a starch hydrolysate or wheat starch
XX hydrolysate which is difficult to filter and to reduce the
XX phospholipid content in an edible oil by treating the oil with
XX phospholipase B to hydrolyse a major portion of the phospholipid
XX and separating an aqueous phase containing the hydrolysed
XX phospholipid from the oil. Phospholipase B is also useful for
XX preparing a dough or a baked product having improved property such
XX as increased strength, elasticity, stability and extensibility. Two
XX primers (AAAS4536, AAAS4537) were used to amplify the phospholipase B
XX genomic DNA sequence.
XX
XX Sequence 27 BP; 8 A; 6 C; 5 G; 8 T; 0 other;

Query Match 0.8%; Score 18; DB 22; Length 27;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 atgaagtcaccgcctctg 528
Db 10 atgaagtcaccgcctctg 27

RESULT 11
AAC61921
ID AAC61921 standard; DNA; 40 BP.
XX
XX AAC61921;

us-09-426-072-1.oli.rng
XX
XX 06-MAR-2001 (first entry)
XX
XX PCR primer for DNA encoding a prenyltransferase designated ATPT3.
XX
XX Prenyltransferase: ATPT1; ATPT2; ATPT3; ATPT4; ATPT5; ATPT6; ATPT7;
XX ATPT8; ATPT9; ATPT10; ATPT11; ATPT12; tocopherol; homogentisic acid;
XX phytylpyrophosphate; 2-methyl-6-phytylbenzoquinol; antioxidant;
XX nutritional supplement; PCR primer; ss.
XX
XX Arabidopsis sp.
XX
XX WO2000063391-A2.
XX
XX 26-OCT-2000.
XX
XX 14-APR-2000; 2000WO-US10368.
XX
XX 15-APR-1999; 99US-0129899.
XX
XX 30-JUL-1999; 99US-0146461.
XX
XX (CALJ ) CALGENE LLC.
XX
XX Savidge B, Lassner MW, Weiss JD, Post-Beittemmiller D;
XX
XX WPI; 2000-647519/62.
XX
XX An isolated nucleic acid sequence encoding prenyltransferase used to
XX transform plant cells to increase the production of tocopherols -
XX
XX Example 2; Page 26; 114pp; English.
XX
XX PCR primers AAC61920-21 were used to amplify DNA encoding a
XX prenyltransferase. The specification describes prenyltransferases
XX designated ATPT1, ATPT2, ATPT3, ATPT4, ATPT5, ATPT6, ATPT7, ATPT8,
XX ATPT9, ATPT10, ATPT11, and ATPT12. The biosynthesis of alpha-tocopherol
XX in higher plants involves the condensation of homogentisic acid and
XX phytylpyrophosphate to form 2-methyl-6-phytylbenzoquinol, which can form
XX various tocopherols. The prenyltransferase polynucleotides are useful in
XX transforming host cells to alter the expression of prenyltransferase in
XX these cells. The transformed cells are used in the production of
XX tocopherols which are of use in the pharmaceutical industry as
XX antioxidants and also in the food industry as nutritional supplements.
XX
XX Sequence 40 BP; 12 A; 9 C; 7 G; 12 T; 0 other;

Query Match 0.8%; Score 18; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1474 ttgaaactcttctcaag 1491
Db 16 ttgaaactcttctcaag 33

RESULT 12
AAH33057
ID AAH33057 standard; cDNA; 1070 BP.
XX
XX AAH33057;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:113.
XX
XX Human colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX

```

PD 05-APR-2001.  
XX 28-SEP-2000; 2000WO-US26524.  
XX 29-SEP-1999; 99US-015137.  
XX 03-NOV-1999; 99US-0163280.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX WPI: 2001-235357/24.  
XX P-PSDB; AAG73626.  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
XX Claim 1; Page 2290-2291; 9803pp; English.  
XX AAG732943 to AAG73195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the patients own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated ps,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAG73196 to AAG73204  
XX and AAG77789 represent sequences used in the exemplification of the  
XX present invention.  
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
XX missing at time of publication, meaning no sequences are present for  
XX SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX Sequence 1070 BP; 237 A; 292 C; 293 G; 237 T; 11 other;  
Query Match 0.8%; Score 18; DB 22; Length 1070;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1644 gacaagaccctgctcttc 1661  
Db 692 gacaagaccctgctcttc 709  
|||||||  
RESULT 13  
AAC61895/c  
ID AAC61895 standard; DNA; 1224 BP.  
XX AAC61895;  
XX 06-MAR-2001 (first entry)  
XX DNA encoding an aromatic class prenyltransferase designated ATPT3.  
XX prenyltransferase: ATPT1; ATPT2; ATPT3; ATPT4; ATPT5; ATPT6; ATPT7;  
XX ATPT8; ATPT9; ATPT10; ATPT11; ATPT12; tocopherol; homogentisic acid;  
XX phytylpyrophosphate; 2-methyl-6phytylbenzoquinol; antioxidant;  
XX nutritional supplement; ss.  
XX Arabidopsis sp.  
XX Key Location/Qualifiers  
FH 1..1224  
FT CDS /\*tag= a  
FT /product= "prenyltransferase ATPT3"  
XX

PN WO200063391-A2.  
XX 26-OCT-2000.  
XX 14-APR-2000; 2000WO-US10368.  
XX 15-APR-1999; 99US-0129899.  
XX 30-JUL-1999; 99US-0146461.  
XX (CALJ ) CALGENE LLC.  
XX Savidge B, Lassner MW, Weiss JD, Post-Beittenmiller D;  
XX WPI: 2000-647519/62.  
XX P-PSDB; AAB19409.  
XX An isolated nucleic acid sequence encoding prenyltransferase used to  
XX transform plant cells to increase the production of tocopherols -  
XX Example 1; Page 76; 114pp; English.  
XX The present sequence encodes a prenyltransferase. The specification  
XX describes prenyltransferases designated ATPT1, ATPT2, ATPT3, ATPT4,  
XX ATPT5, ATPT6, ATPT7, ATPT8, ATPT9, ATPT10, ATPT11, and ATPT12. The  
XX biosynthesis of alpha-tocopherol in higher plants involves the  
XX condensation of homogentisic acid and phytylpyrophosphate to form  
XX 2-methyl-6phytylbenzoquinol, which can form various tocopherols. The  
XX prenyltransferase polynucleotides are useful in transforming host cells  
XX to alter the expression of prenyltransferase in these cells. The  
XX transformed cells are used in the production of tocopherols which are  
XX of use in the pharmaceutical industry as antioxidants and also in the  
XX food industry as nutritional supplements.  
XX Sequence 1224 BP; 308 A; 220 C; 288 G; 408 T; 0 other;  
Query Match 0.8%; Score 18; DB 21; Length 1224;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1474 ttgaaaaactcttccaag 1491  
Db 1221 TTGAAAACCTTCTTCCAG 1204  
|||||||  
RESULT 14  
AAZ29801  
ID AAZ29801 standard; DNA; 1307 BP.  
XX AAZ29801;  
XX 27-MAR-2000 (first entry)  
XX Sequence downstream to Ds insert in tomato mutant UQ13.  
XX Tomato mutant UQ13; UDP-glucose-pyrophosphorylase gene; expansin gene;  
XX PMGS; phenotype modifying genetic sequence; plant pathogen resistance;  
XX senescence timing; starch metabolism; ds.  
XX Lycopersicon esculentum.  
XX WO9963068-A1.  
XX 09-DEC-1999.  
XX 04-JUN-1999; 99WO-AU00434.  
XX 04-JUN-1998; 98AU-0003901.  
XX 04-JUN-1998; 98AU-0003903.  
XX 25-SEP-1998; 98AU-0006169.  
XX 25-SEP-1998; 98AU-0006174.  
XX (UYQU ) UNIV QUEENSLAND.  
PA



XX Carroll BJ;  
XX WPI; 2000-116368/10.  
XX  
XX New polynucleotides that increase gene expression in plants used to  
XX produce transgenic plants with resistance to plant pathogens -  
XX  
XX Disclosure: Page 67; 93pp; English.  
XX  
XX The present sequence is the downstream portion of Ds insert containing  
XX PMS in tomato mutant UO13. The Ds insertion is associated with a  
XX putative UDP-glucose-pyrophosphorylase and/or expansin gene, which are  
XX potentially involved in starch metabolism. This resulted in nos: BAR  
XX expression. This can be used to produce transgenic plants having altered  
XX phenotypic traits, such as resistance to plant pathogens, senescence  
XX timing, starch metabolism, cell growth, expansion and/or division, and  
XX the shape of cells, tissues or organs.  
XX  
XX Sequence 1307 BP; 497 A; 163 C; 170 G; 471 T; 6 other;  
XX  
XX  
XX Query Match 0.8%; Score 18; DB 21; Length 1307;  
XX Best Local Similarity 100.0%; Pred. No. 58;  
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1212 ctgtatatttttaatttt 1229  
XX |  
XX Db 780 ctgtatatttttaatttt 797  
XX  
XX  
XX RESULT 15  
XX AAI62830  
XX ID AAI62830 standard; DNA; 1394 BP.  
XX AC AAI62830;  
XX  
XX 22-OCT-2001 (first entry)  
XX  
XX Human genomic DNA SEQ ID NO 158.  
XX  
XX Human; neotropic; neuroprotective; cytostatic; dermatological; virucide;  
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnary;  
XX antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;  
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;  
XX ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155449-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01346.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUL-2000; 2000US-0216880.  
XX 14-JUL-2000; 2000US-0218290.  
XX 14-AUG-2000; 2000US-0225447.  
XX 01-SEP-2000; 2000US-0229343.  
XX 06-SEP-2000; 2000US-0230437.  
XX 08-SEP-2000; 2000US-0231243.  
XX 25-SEP-2000; 2000US-0234997.  
XX 29-SEP-2000; 2000US-0236367.  
XX 13-OCT-2000; 2000US-0239837.  
XX 08-NOV-2000; 2000US-0246476.  
XX 08-NOV-2000; 2000US-0246477.  
XX 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246528.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249265.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-476225/51.  
XX  
XX Novel plasma membrane associated proteins useful for diagnosing,  
XX treating, preventing and/or prognosing disorders related to the  
XX proteins, including cancer, immune response and neuronal disorders -  
XX  
XX Example 2; SEQ ID NO 158; 532pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (AAI62752-AAI62961) and proteins  
XX (AAM42347-AAM42415) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1394 BP; 313 A; 385 C; 338 G; 358 T; 0 other;  
XX  
XX  
XX Query Match 0.8%; Score 18; DB 22; Length 1394;  
XX Best Local Similarity 100.0%; Pred. No. 58;  
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1213 tgtatatttttaattttg 1230  
XX |  
XX Db 1340 tgtatatttttaattttg 1357  
XX  
XX  
XX Search completed: December 4, 2001, 18:37:46  
XX Job time: 5639 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 15:38:16 ; Search time 3277.91 Seconds  
(without alignments)  
11666.111 Million cell updates/sec

Title: US-09-426-072-1

Perfect score: 2318

Sequence: 1 aattcttggaacaatccc.....gcaccccgatacattaatt 2318

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_ph.\*

7: gb\_pl.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pi.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2318	100.0	2318	6	AR141633	Sequence
2	2318	100.0	2318	6	AX127521	Sequence
c 3	22	0.9	10807	1	AE004899	Pseudomon
c 4	22	0.9	170970	2	AC021564	Homo sapi
c 5	21	0.9	30	6	AR141635	Sequence
c 6	21	0.9	30	6	AX127524	Sequence
c 7	21	0.9	38627	9	AC011538	Homo sapi
c 8	21	0.9	40611	9	AC021090	Homo sapi
c 9	21	0.9	105538	2	AC092279	Homo sapi
c 10	21	0.9	153173	2	AC091983	Homo sapi
c 11	21	0.9	184606	2	AC072046	Homo sapi
c 12	21	0.9	200889	2	AC067865	Homo sapi
c 13	21	0.9	201312	2	AC067900	Homo sapi
c 14	21	0.9	209910	2	AC073534	Homo sapi
c 15	20	0.9	240	8	CNS01BDF	Botrytis
c 16	20	0.9	636	8	CNS01CHX	Botrytis
c 17	20	0.9	2215	3	S39048	knob associ
c 18	20	0.9	2521	9	BC007685	Homo sapi
c 19	20	0.9	3748	1	AF269548	Staphyloc
c 20	20	0.9	3748	6	AX144868	Sequence
c 21	20	0.9	10645	1	AE008335	Agrobacte
c 22	20	0.9	11093	1	AE005134	Halobacte
c 23	20	0.9	12029	3	AE001372	Plasmodi
c 24	20	0.9	63385	9	HSJ474G15	Human DNA
c 25	20	0.9	80765	9	AC004009	Homo sapi
c 26	20	0.9	102393	8	AC000107	Genomic s
c 27	20	0.9	114364	9	AC011666	Human Chr
c 28	20	0.9	126228	9	AL513211	Human DNA
c 29	20	0.9	167147	2	AC026860	Homo sapi
c 30	20	0.9	170594	2	AP003385	Homo sapi
c 31	20	0.9	170956	2	AC024973	Homo sapi
c 32	20	0.9	172805	9	AC066608	Homo sapi
c 33	20	0.9	175084	2	AL161636	Homo sapi
c 34	20	0.9	175233	2	AC022941	Homo sapi
c 35	20	0.9	202760	2	AC021987	Homo sapi
c 36	20	0.9	202924	2	AC079240	Homo sapi
c 37	20	0.9	203257	9	AC013737	Homo sapi
c 38	20	0.9	212494	2	AC074037	Mus muscu
c 39	19	0.8	287	10	AF045954	Sequence
c 40	19	0.8	304	6	AX014844	Sequence
c 41	19	0.8	1138	9	HSA012584	Homo sapi
c 42	19	0.8	1231	9	BC000805	Homo sapi
c 43	19	0.8	1260	1	AF314196	Pseudomon
c 44	19	0.8	1263	6	AX025011	Sequence
c 45	19	0.8	1281	6	AX025012	Sequence

#### ALIGNMENTS

RESULT 1	AR141633	2318 bp	DNA	PAT	08-AUG-2001
LOCUS	Sequence 1 from patent US 6146869.				
DEFINITION	AR141633				
ACCESSION	AR141633				
VERSION	AR141633.1	GI:15101149			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 2318)				
AUTHORS	Harris, P. and Brown, K.M.				
TITLE	Polypeptides having phospholipase B activity and nucleic acids encoding same				
JOURNAL	Patent: US 6146869-A 1 14-NOV-2000;				
FEATURES	Location/Qualifiers				
source	1..2318				
BASE COUNT	554 a 655 c 534 g 575 t				
ORIGIN	/organism="unknown"				

Query Match					100.0%; Score 2318; DB 6; Length 2318;				
Best Local Similarity					100.0%; Pred. No. 0;				
Matches 2318; Conservative					0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	aattcctgacacaatcccttgaagatgataagggatgaactgagatgctaag	60						
Db	1	AAATTCCTGGACAAATCCCTTTGAGAGTGATAGTAAAGGGGATGAACCTGAGATGCTAAG	60						
Qy	61	ctcaagcttcctgccttggcgttagcgtgcccccatgcttgggttagtgcttctctccacc	120						
Db	61	CTCACGTTCTTGGCCTTGGCGTAGGCTGCCCATGCTTGGTTAGTGTCTCTCCACC	120						
Qy	121	ctagtagcttagccattgcttccaatccccattccccctcgcacacacacttatattat	180						
Db	121	CTAGTAGCTTAGCCATTGCTTGTCCAATFCCCATTCCTCCATCTCCGATCAGCTCTATTAT	180						
Qy	181	gaccaagcgtgcaaatgagagccccactcatggtggcctccacttcattgttttcttt	240						
Db	181	GACCAAGGCTGTCAAAATGAGAGCCCACTCATGTGGCTCCACTTTCATATATTTCTTTT	240						
Qy	241	gtcgataagactgtcatcccgctccggtggcttcacagaagaacgatttgaagatg	300						
Db	241	GTGATAGAAGCTGTCATCCCGTCTCCGTTGGCTTCCACGAAGAAGATTGTAAAGATGG	300						
Qy	301	atagcgataggctgccccttggctggtatgaagccgaagtccctcgtcctcgatgat	360						
Db	301	ATAGCGATAGGCTGCCCTTGGTGGTATGAAGCCGAAGTCCCTCGCTCGATGATAT	360						
Qy	361	ggccttcaaccgacacacagcagcagtggtatccgaatggagcgaagagtgactcttc	420						
Db	361	GCCTTCAACCCGACGACACGACGCTGCTCCGTTGGCTTCCACGAAGAAGATTGTAAAGATGG	420						
Qy	421	caagtgggctaataagccttccctggcgtggtcagagtcacagtcacagtcacagtcac	480						
Db	421	CAGTGGGCTATATAAGCCTTCCCTGGCCGGGCTCAGACTCCCTAGCAAGCAGCATGTT	480						
Qy	481	caagctcagcagcctcaatccgagagtaaacatgaagtcacacgctccttactgcttc	540						
Db	481	CAAGCTCAGCAGCTCAATCCGGAGTAACATGAAGTCCACCGCTCTCTTACTGGTCTC	540						
Qy	541	ggccttggcctctctcgtctctagcctgctcagtcagtcagtcagtcagtcagtcagtc	600						
Db	541	GGCCTTGGCCTCTCTCGCTCTCTGCTTGTCTAGCCCTGTACGCTCCAGTATACGAGCGTGA	600						
Qy	601	gaagcctcttcgatacaagcctggctcccaagagtgccattgagaacttgaagacag	660						
Db	601	GAAGCCCTTTCGGATACAAAGCTGGCTCCAAAGGAGTCCATTGAGAACTTGAAGGACAAG	660						
Qy	661	gtcgaagaacattgtctggcttatctcgagaacagggttggtgccccttaagatattcat	720						
Db	661	GTGAGAAACATTGTGCTGCTTATCTCGAAGACAGGTTGTGTGCCCTTACGATATTTCAT	720						
Qy	721	atgtggaataataatccctcaattcagcttggatgtggaagcagcagcactaaacatat	780						
Db	721	ATGTGGAATAATAATCCCTCAATTCAGCTTGGTATGATGGAAGCAGCAGCACTAAACAT	780						
Qy	781	ggctcaagccttcgataaacttcctggaggcgtgcccgaaggaatggaaaccca	840						
Db	781	GGTCCAGATCTTTCGATTAACATTTCTGGGAGGCGTGCCTGCCCAAGGACTGGACAAACCGA	840						
Qy	841	tcaacaacgcccgttctgcaactcaagaatgagagcaccatcctcgggcaagtaact	900						
Db	841	TCAACAACGGCCGTTCTGCAACTTCAAGAATGGGAGCGACCACTCCCTGGGCAAGTACT	900						
Qy	901	gtactcaggccaaggaactatgttcggtttcaacagatcaagaccacccctgactggta	960						
Db	901	GTACTCAGGCCAAGGACTATGATTTCCGTGTTCAACAGATCCAGACCCTCCGTGACTGGTA	960						
Qy	961	ataactgggttctacggaaacttacaccccaacaaatggtgcatgtgccagtggaag	1020						
Db	961	ATAACTTGTGAGTTCTACGGAACCTTACACCCCAAAACAAATGGTGGATTGCCAGTGCAAGG	1020						

Qy	1021	tcgtccgacacagtcctggttctcaagcacagcttaacgactcaccacaaactggccc	1080						
Db	1021	TCGTCGCCGACCACTGCTGCTTCTCAACGACAGCTTAACGACTACCCCAAACTGGCCC	1080						
Qy	1081	cagaagggcgaacaggcaagtgtatggatatactatacagggagggaggttctcactgcg	1140						
Db	1081	CAGAAGGGCGACAAAGGCAAGTGTATACGGATACCTATACGGAGGAGGTTCTCTAGCTCG	1140						
Qy	1141	tggaaccttggaatgaattcactacttcaacagctgattctcgtgttctcgtgcgtaa	1200						
Db	1141	TGGACCTTGTGGATGATGTTCTACTTTCACAGCTGGTTCGTGTGTTCTTGGGGTAA	1200						
Qy	1201	gtcgataaactctgtatatatttttgaactaacgcttcaaaaagcctcaccacccca	1260						
Db	1201	GTGCGATAAATCTGTATATTTTAAATTTGACTTAACGGCTTCAAAAGCCTACCAACCCCA	1260						
Qy	1261	accgcttgctgctctggcgaagaccgctgctgggca tggcaagaatgacatgacttcc	1320						
Db	1261	ACCGCTTGTGGCTCTGGCAGAACCGCTGCTGGGCATGGCAAGANTGAGATGACTTCC	1320						
Qy	1321	tgaactatgggtatctctagcaagtcctatctcgagggccgcaacgagaagggcgtgctc	1380						
Db	1321	TGAACATATGTTATCTCTAGCAAGTCCATCTTCGAGGCCGCCAACGAGAGGGCTGCTCT	1380						
Qy	1381	ggctcaactacgatggcaccacaggaatctgaacacgattctctcttccactacg	1440						
Db	1381	GGCTCAACTACGATGGGACCAACGAGAAATTCGAACCGGATTCCTCTTCTTCCACTAG	1440						
Qy	1441	tcaaccagacacctcccggtccaaagctgggtgcccgttgaaactcttccaagacgctacc	1500						
Db	1441	TCAACCAGACCTCCGCTCCAAAGCTGGTGGCTGTTGAAATCTTTCAAAGACGCTTACC	1500						
Qy	1501	tcggtctcctccctaaattctctacattaacccctcctcgtcgtcgccaccaacacac	1560						
Db	1501	TCGGTCTCTCCCTTAAATTTCTTACATTAAACCTCTCTGTCGGGCAACCAACCAACT	1560						
Qy	1561	ccatgacccccacggtaacgtctctactaggtgaggtctctcgaacagatctatgatg	1620						
Db	1561	CCATGCAACCCACCGGTAAAGCTCTCTACGGTGGGTCTTCGTCGAAGCAGATCTATGATG	1620						
Qy	1621	ccattcgccagggcctcagtgagggaagacctgctctcaattacactacgacgagacg	1680						
Db	1621	CAATTGCGCAGGGCCCTCAGTGGGACAGACCTGCTCTTCAATTACCTTACGACGACCG	1680						
Qy	1681	gtggcttctacgacctgtccctccctcctcgcgtccgcccggagaaacctgaactaca	1740						
Db	1681	GTGGCTTCTACGACCAATGTCCCTCCCTCTCGCCCTCCGCGGACAACTGACCTACA	1740						
Qy	1741	ctgagactggaagaaacggctcagaataacactcttcaacttcgacctcgtgggtggccga	1800						
Db	1741	CTGAGACTTGGCAAGAGGGTCAAGAAATACACTCTTCACTTCGACCGTCTGGGTGGCCGA	1800						
Qy	1801	tgccagctgggtttatctcccttcaagaaggggatacatcgagcagtcaggaacgg	1860						
Db	1801	TGCCAGCTTGGGTATCTCCCTTACAGTAAGAGGGATACATCAGCAGTACGGAACGG	1860						
Qy	1861	atcccgctcagcggcaagcccgctccctacagtgctacactccgtcctcgaactctcgat	1920						
Db	1861	ATCCCGTCAACGGCAAGCCCGCTCCCTACAGTGTACCTCCGCTCCTCAAGACTCTCGAT	1920						
Qy	1921	atctctgggacaactcgaggaactcaacctcgtgtcggcccaactctcactcttgcatacc	1980						
Db	1921	ATCTTGGGACATCGAGGACTTCAACCCCTGCTGTGGCCCACTCTCTCATCTTTGATACCC	1980						
Qy	1981	tgatcggcacgactttgctgaggaatgctcctattgtctcgaactccccca tacccttt	2040						
Db	1981	TGATCGGCACGACTTTTGGGTGAGGATGCTCTCTATTGTCTCTCAAGACTCCCCATACCTTT	2040						
Qy	2041	cggataagctcagtcggtcgaggaacagatataaagatgaatgaacgcttctgac	2100						
Db	2041	CGGTATAAGTCTCAGTCCGTCAGTGAGACAGAGGATTAAGAATGATGATGAACGCTTCTGAC	2100						

QY 2101 ttcaatgaacgattacattgttatagagcattgttttgcattatagctacgcttagagcgag 2160  
 |||||  
 Db 2101 TTCACTGAACGATTACATTATAGAGCATTGTTTGTCTATAGCTACGCTAGAGCGAG 2160  
 |||||  
 QY 2161 cgcgatgataagaataaagctgggttatttctctattgtatttatttatttatttatttattt 2220  
 |||||  
 Db 2161 CCGGATGATGATAAGATAAAGCTGGGTATTCTCTATTGTATATTATTATTATTATTATT 2220  
 |||||  
 QY 2221 ttgtataacattggatttataaaggaatttttcttgcacaaactcaactcacggaacagg 2280  
 |||||  
 Db 2221 TTTGATAACATGGGATTTAAAAGGAAATGTTTCTTGCAATCAACTCACGGAACAGG 2280  
 |||||  
 QY 2281 cgttaacttiacgatgactcaccccggtacataaatt 2318  
 |||||  
 Db 2281 CGTAACCTTAGGATGACTGCACCCCGGATACATAATT 2318  
 |||||

## RESULT 2

AX127521 2318 bp DNA PAT 15-MAY-2001  
 DEFINITION Sequence 1 from Patent WO0129222.  
 ACCESSION AX127521  
 VERSION AX127521.1 GI:14134227  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Aspergillus oryzae.  
 Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE  
 AUTHORS  
 TITLE  
 Polypeptides having phospholipase b activity and nucleic acids  
 encoding same

JOURNAL  
 Patent: WO 0129222-A 1 26-APR-2001;

NOVO NORDISK BIOTECH, INC. (US)

Location/Qualifiers

1. .2318

/organism="Aspergillus oryzae"

/db\_xref="taxon:5062"

BASE COUNT 554 a 655 c 534 g 575 t

ORIGIN

Query Match 100.0%; Score 2318; DB 6; Length 2318;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aattcctggagcaacaatcccttggagagtgatagtaaggagggaatgaactgagatgctaag 60  
 |||||  
 Db 1 AATTCTGGGACAAATCCCTTTGAGAGTGATAGTAAGGGGGGATGAATCTGAGATGCTAAG 60  
 |||||  
 QY 61 ctcaagcttctgccttggcgtaggtgccccatgtcttggtttagtgccttctctccacc 120  
 |||||  
 Db 61 CTCACGTTCTTCGCTTGGCGTAGGCTGCCCATGTCTTGGTTTGTGCTTCTCTCCACC 120  
 |||||  
 QY 121 ctatgattgagcattgcttgcacatccatccatccatccatccatccatccatccatccat 180  
 |||||  
 Db 121 CTAGTAGCTTAGCATTGCTGTGTCATTCATTCATTCATTCATTCATTCATTCATTCAT 180  
 |||||  
 QY 181 gacaaaggctgcaaatgagagcccccaactcatgtggcctccacttcatttattgtttcttt 240  
 |||||  
 Db 181 GACCAAGGCTGTCAAAATGAGAGCGCCCACTCATGTGGCTCCACTTCATATTGTTTCTTT 240  
 |||||  
 QY 241 gtcgataagctgctatccgtctccgtgcttccaggaagcgaatttgaagatgg 300  
 |||||  
 Db 241 GTCGATAAGCTGTATCCCGTCTCCGGTGGCTTCCACGAAGAACGATTGTTGAAGATGG 300  
 |||||  
 QY 301 atagcgataggctgaccttggctggtatgaagcggaagtccctcgctgcgtcgatggat 360  
 |||||  
 Db 301 ATAGCGATAGGCTGCCCTTGGTGGTATGAAGCGCAAGTCCCTCGCTCGCTCGATGGAT 360  
 |||||  
 QY 361 gccttcaaccgacgacagcgctggtatccgaatggagcgaaggttagctcttc 420  
 |||||  
 Db 361 GCGCTTCAACCCGACGACAGCGACGTGCTATCGGANTGGGACGGAAGTAGCTCTCTC 420  
 |||||

QY 421 cagtgggctataaagccttccctggccgggctcagactccctagcaagcagcattgtt 480  
 |||||  
 Db 421 CAGTGGGCTATATAAGCCTTTCCCTGGCCGGGCTCAGACTCCCTAGCAAGCAGCATTTGT 480  
 |||||  
 QY 481 caagctcacacacctcaatccgagagtaaacatgaagtcacacgctctgcttacttgcttc 540  
 |||||  
 Db 481 CAAGCTCACGACCTCAATCCCGGAGTAAACATGAAGTCCACCGCTCTGCTTACTGGTCTC 540  
 |||||  
 QY 541 ggcctcttggcctctcctggtcttgcctagcctgtcaacgtccgaggtatagcagcgtgca 600  
 |||||  
 Db 541 GGCCTCTTGGCCTCTCTCGGCTTCTAGCCCTGTCCACGCTCCGAGTATAGCAGCGTGGCA 600  
 |||||  
 QY 601 aaagcccttccgatacaagccttgctccaaagagtcatttgagaacttgaaggaacaag 660  
 |||||  
 Db 601 GAAGCCCTTTTCGGATACAAGCTTGGCTCCNAGGAGTCCATTGAGAACTTGAAGGACAA 660  
 |||||  
 QY 661 gtcgagaacattgtctggcttattctcgagaacaggttggcttgccttaacgatatcat 720  
 |||||  
 Db 661 GTCGAGAACATTGTCTGGCTTATTCTCGAGAACAGGTTTGTGTCCTTACGATATTCTAT 720  
 |||||  
 QY 721 atgtggaataaataatctcaattcagcttgggtatgtgaagcagcagcactaatacat 780  
 |||||  
 Db 721 ATGTGGAATAATAATTCCTCAATTACGCTTGTGTTATGTGAAGACGAGCACTAAACATAT 780  
 |||||  
 QY 781 ggtccagatccttcgataacattctggaggcggtgctgcccgaagagcagcagcagcag 840  
 |||||  
 Db 781 GGTCCAGATCCTTCGATAACATTCTGGAGGCGGTGGCGCGCCCAAGAGCTGACAAACCGA 840  
 |||||  
 QY 841 tcaacaacgcccgtctcactcaactcaagaatcgagcagcagcagcagcagcagcagcag 900  
 |||||  
 Db 841 TCAACAACGGCCGCTTCTGCNACTACAGANTGCGAGGAGCCCACTCTCGGGCAAGTACT 900  
 |||||  
 QY 901 gtactcaggccaaggactatgtatccgtgttcaacagatccagacacacacacacacacac 960  
 |||||  
 Db 901 GTACTCAGGCCAAGGACTATGATTTCCGTGTTCAACGATCCAGACACACTCCCGTAGGTA 960  
 |||||  
 QY 961 ataacttgagttctcaggaacttacacccaacaatggtgcgattgcccagtggaag 1020  
 |||||  
 Db 961 ATAACTTGAGTTCTACGGAACTTACACCCCAACAATGTGCGATTGCCAGTGGCAAGG 1020  
 |||||  
 QY 1021 tgcctgcgacacagcttgccttctcactcaacgcagcagcagcagcagcagcagcagcag 1080  
 |||||  
 Db 1021 TCGTCGCGACACAGCTGCGCTTCTCAACGCACAGCTTAACAGACTACCCCAAACTGGCCC 1080  
 |||||  
 QY 1081 cagaagagcgacaggaagcaagtgatgggatactatacggaggaggaggttctcactgcctg 1140  
 |||||  
 Db 1081 CAGAGAGCGCACAAAGCAAGTATGGGATACTATACGGAGGAGGAGGTCTCTACGCTCG 1140  
 |||||  
 QY 1141 tggaccttggatggatgactcaacttcaacagcagctgttctcggtgtgttctctggcgtaa 1200  
 |||||  
 Db 1141 TGGACCTTGTGGATGAGTTTCACTTACTTCAACAGCTGTTTCTGCTGTGTTCTTGGCGTAA 1200  
 |||||  
 QY 1201 gtcgataaaactctgataatcttcttcttcttcttcttcttcttcttcttcttcttct 1260  
 |||||  
 Db 1201 GTCGATAAACTCTGATATATTTTAAATTTTGAATTAACGCGTTCAAAAGCTTACCAACCCA 1260  
 |||||  
 QY 1261 accgcttggcgtctctggcaggaacccgtgctgggcatggcgaagaatgaagcagcagcag 1320  
 |||||  
 Db 1261 ACCGCTTGGCGTCTTGGCAGGAACCGCTGCTGGGCAATGGCAAGATGACGATGACTTCC 1320  
 |||||  
 QY 1321 tgaactatgttctctctagcaagtcctcttcgagggcgccgaacgaggaagggcggttct 1380  
 |||||  
 Db 1321 TGAACATATGTTATCTCTAGCAAGTCCATCTTCGAGCGCGCCCAACGAGAGGCGGTGCTCT 1380  
 |||||  
 QY 1381 ggcctcaactcagatggcagcaacgagagattcgaacgggattctctcttctcctcctacg 1440  
 |||||  
 Db 1381 GGCCTCAACTCAGATGGCAGCAACGAGAAATTCGAACCGGATTTCTCTTCTTCTACCTACG 1440  
 |||||  
 QY 1441 tcaaccagacacctcccggtcccaacggtggtcccggttgaaaaactcttccaaagacgctacc 1500  
 |||||  
 Db 1441 TCAACCAAGACCTCCCGGTCCAAGCTGCTGCCGTTGAAACCTTCTTCCAAGACGCTTACC 1500  
 |||||

QY 1501 tcggtgtctccctaaattctcttattacataaacccctctctgctcgccaccacccaact 1560  
|||||  
Db 1501 TCGGTGTCCTCCCTAAATTCCTTACATTAACCCCTCTCTGCTCGCGCACCAACCAACT 1560  
QY 1561 ccattgacccccaccggaacgtctctctacgtgaggtctctctcaagcagatctatgatg 1620  
|||||  
Db 1561 CCATGACCCACCAGGTAAGCTCTCTACGGTGAGGTCTCGTCAAGCAGATCTATGATG 1620  
QY 1621 ccattgacccccaccggtcagtgaggaagaccctctctctctctctctctctctctct 1680  
|||||  
Db 1621 CCATTCGCCAGGGCCCTCAGTGGGACAAAGACCCCTCTCTCTCTCTCTCTCTCTCTCT 1680  
QY 1681 gtggtctctacgacatgtccctccctctctcgctcgcccgccgagacacactgacctaca 1740  
|||||  
Db 1681 GTGGCTCTACGACCATGTCCTCCCTCTCGCGTCTCGCGTCCCGCGGAACCTGACCTACA 1740  
QY 1741 ctgagactcggaagacggtcagaataacacactctctctctctctctctctctctctct 1800  
|||||  
Db 1741 CTGAGACTCGGAAGACGGTCAGAAATACACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1800  
QY 1801 tgcgacctgggtlatctccctctctacagtaagaggatatactcagcagtcagcaacgg 1860  
|||||  
Db 1801 TGGCGACCTGGGTATCTCCCTCTACAGTAAGAGGATACATCGAGCAGTACGGAACGG 1860  
QY 1861 atccctcaagggacgctccctctctctctctctctctctctctctctctctctctct 1920  
|||||  
Db 1861 ATCCCGTCAAGGGACGGCTCCCTCTACAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920  
QY 1921 atctctgggacatcgagactctcaccctctgtcgccactctcctctctctctctct 1980  
|||||  
Db 1921 ATCTCTGGGACATCGAGACTTCAACCTCTGTCTCGCCACTCTCCATCTTTTCGATCACC 1980  
QY 1981 tgcagcagcacttctgagagatgctctctctctctctctctctctctctctctctctct 2040  
|||||  
Db 1981 TGATCGGACGACTTTGGTGAGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2040  
QY 2041 cggtaagctcagtcgctgagcagagattaaaagtgatgaacgctctctctctctctct 2100  
|||||  
Db 2041 CGGTATAAGTCTCAGTCGGTGCAGTGAGAGAGATTAAGATGATGATGAACGTTCTGAC 2100  
QY 2101 ttcagtgacgattacatgattatagagatgtttctctctctctctctctctctctct 2160  
|||||  
Db 2101 TTCTGTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
QY 2161 cggatgataagataagcgtgggtattctctctctctctctctctctctctctctctctct 2220  
|||||  
Db 2161 CGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
QY 2221 ttgataacatggattataaaggaatgtttctctctctctctctctctctctctctct 2280  
|||||  
Db 2221 TTTGATAACATGGGATTTAAAGGAATGTTTCTTTCGACATCACTCACTCACTCACTCA 2280  
QY 2281 cgtactttacgatacgtgaccccgagatacattatt 2318  
|||||  
Db 2281 CGTAACCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2318

RESULT 3  
AE004899/c  
LOCUS  
DEFINITION  
Pseudomonas aeruginosa PA01, section 460 of 529 of the complete genome.  
ACCESSION  
AE004899 AE004091  
VERSION  
AE004899.1 GI:9951126  
KEYWORDS  
Pseudomonas aeruginosa.  
SOURCE  
Pseudomonas aeruginosa.  
ORGANISM  
Bacteria; Proteobacteria; gamma subdivision, Pseudomonadaceae;  
Pseudomonas.  
REFERENCE  
1. (bases 1 to 10807)  
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Huftagie,W.O., Kowalik,D.J., Lagrou,M., Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Paulsen,I.T., Reizer,J., Salier,M.H., Wong,G.K., Wu,Z., Paulsen,I.T., Lory,S. and Olson,M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University Of Washington,  
Box 352145, Seattle, WA 98195, USA  
Location/Qualifiers  
1. .10807  
/organism="Pseudomonas aeruginosa"  
/strain="PA01"  
/db\_xref="taxon:287"  
83..1372  
/gene="purD"  
/note="PA4856"  
83..1372  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosylamine--glycine ligase"  
/protein\_id="AAG08240.1"  
/db\_xref="GI:9951127"  
/translation="MNVLIIGSGREHALWVADPRVAKVFPVAGNAGTAEAKCE  
NVALDVALEQLADPAKKNVLTIVGPEAPLVAGVDLPERGLDIFGPTAGAAQLG  
SKAFTDLARHRIPTAAYRNTEVEPALVLEHQQGAPVIRADGLAGKGVYVNNL  
DEAAAVRMLAGNAGFADGASRVVIEEFDGGEASFIVVDGQNVLPWATSDHRRVG  
DGGPNTGGMGAYSPAPVTVNEHVLDIVTYPTVMAEENGVNVPATSGHRLMD  
KAGPKVTEFCRDEPDTQINVRLSESLVLEAAALAKDKVATWTFRTVGVV  
LSAGGPGDYAKGEVTEGLAAALDKGVFHAGTALDKGVVTSGRVLCATAGESV  
SAAQQOAVKLAEKIRWNGCFYKIDGIRAIARERGES"  
1476..4304  
/gene="PA4856"  
1476..4304  
/gene="PA4856"  
/codon\_start=1  
/transl\_table=11  
/product="probable sensor/response regulator hybrid"  
/protein\_id="AAG08241.1"  
/db\_xref="GI:9951128"  
/translation="MVRIRIAGLIVSFLILLGLPMSFVADGAVSSVLTQTATTPT  
SANWRLLRPSAQLRIDVLRQKQFPLAKRSFTFPASPOAVLQVLPQAKVPS  
WLRTPAPRVQVLDYLVQDQVLDQHTGESRPPQERPLPSRYLSPVQKPMWLY  
VRMTSNHPLMWFQIDBAGLVGLKXPAYAFGMLLGGMLLLNLYNLRFPASRSASL  
WLAARHAALVCAANLGLVAFWPLGLAFNGLSTADLGLAAVLSLLAFACSPFRGTA  
ESRLNLQGEALILAVGAIATAFTQOLWFWNLVLLVLSLSVPLTAHWHVFRGTA  
PARLIVAGMIVFNAGFMVFLPVTGKQDPLVGLVGFATLGLVLSVLTQROR  
LIQQLNLOORTSEAAHTAEQKAEFLAKISHEIRTPMGVIGMTLELLIGTPLSAKQ  
DYVQTIHSGNELLTINELIDISKLESQLEDEVDFDNLALIEDCIDIFRVAKQOR  
RIELISFTQPVPRVIGGDTPLRQVRLSLDQNAFKOTEGEILLVVALDQGETPR  
RIAVDGGHPDPAKEREALAEHSDPLSKLSGLSHGLIITAROLVRLMGGEFGIQ  
SGSQGCTLTSLPLDQOLENTADLPGLOGARLLVDDNETCKVLVQCCSGWGL  
NVSASVSGREALAQLRTKAHLREYFDVLLDQMPGTGQLAALKIKEPDLNHDILL  
IMLTGISNAPSRIARNAGIKRLAKPVGATYTLKATLADLAORGVSQVGTQFDVLV  
AQAPLSDFRILVAEDNSISTKVTGMLNKLINLOPDTASNGEALSAMKATQVLDVL  
MDCMEVLDGFSATQEAWEHQRPPTPVVALTATLHSEKREARLVGMDGHMAKP  
VELSQLRLIAYVWGERRRRQDGLPS"  
4760..5353  
/gene="PA4857"  
4760..5353  
/gene="PA4857"  
/codon\_start=1  
/transl\_table=11

Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,  
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.  
Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
20437337  
2 (bases 1 to 10807)  
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Huftagie,W.O., Kowalik,D.J.,  
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,  
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
Folger,K.R., Kas,A., Paulsen,I.T., Reizer,J., Salier,M.H.,  
Wong,G.K., Wu,Z., Paulsen,I.T., Lory,S. and Olson,M.V.  
Hancock,R.E.W., Lory,S. and Olson,M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University Of Washington,  
Box 352145, Seattle, WA 98195, USA  
Location/Qualifiers  
1. .10807  
/organism="Pseudomonas aeruginosa"  
/strain="PA01"  
/db\_xref="taxon:287"  
83..1372  
/gene="purD"  
/note="PA4856"  
83..1372  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosylamine--glycine ligase"  
/protein\_id="AAG08240.1"  
/db\_xref="GI:9951127"  
/translation="MNVLIIGSGREHALWVADPRVAKVFPVAGNAGTAEAKCE  
NVALDVALEQLADPAKKNVLTIVGPEAPLVAGVDLPERGLDIFGPTAGAAQLG  
SKAFTDLARHRIPTAAYRNTEVEPALVLEHQQGAPVIRADGLAGKGVYVNNL  
DEAAAVRMLAGNAGFADGASRVVIEEFDGGEASFIVVDGQNVLPWATSDHRRVG  
DGGPNTGGMGAYSPAPVTVNEHVLDIVTYPTVMAEENGVNVPATSGHRLMD  
KAGPKVTEFCRDEPDTQINVRLSESLVLEAAALAKDKVATWTFRTVGVV  
LSAGGPGDYAKGEVTEGLAAALDKGVFHAGTALDKGVVTSGRVLCATAGESV  
SAAQQOAVKLAEKIRWNGCFYKIDGIRAIARERGES"  
1476..4304  
/gene="PA4856"  
1476..4304  
/gene="PA4856"  
/codon\_start=1  
/transl\_table=11  
/product="probable sensor/response regulator hybrid"  
/protein\_id="AAG08241.1"  
/db\_xref="GI:9951128"  
/translation="MVRIRIAGLIVSFLILLGLPMSFVADGAVSSVLTQTATTPT  
SANWRLLRPSAQLRIDVLRQKQFPLAKRSFTFPASPOAVLQVLPQAKVPS  
WLRTPAPRVQVLDYLVQDQVLDQHTGESRPPQERPLPSRYLSPVQKPMWLY  
VRMTSNHPLMWFQIDBAGLVGLKXPAYAFGMLLGGMLLLNLYNLRFPASRSASL  
WLAARHAALVCAANLGLVAFWPLGLAFNGLSTADLGLAAVLSLLAFACSPFRGTA  
ESRLNLQGEALILAVGAIATAFTQOLWFWNLVLLVLSLSVPLTAHWHVFRGTA  
PARLIVAGMIVFNAGFMVFLPVTGKQDPLVGLVGFATLGLVLSVLTQROR  
LIQQLNLOORTSEAAHTAEQKAEFLAKISHEIRTPMGVIGMTLELLIGTPLSAKQ  
DYVQTIHSGNELLTINELIDISKLESQLEDEVDFDNLALIEDCIDIFRVAKQOR  
RIELISFTQPVPRVIGGDTPLRQVRLSLDQNAFKOTEGEILLVVALDQGETPR  
RIAVDGGHPDPAKEREALAEHSDPLSKLSGLSHGLIITAROLVRLMGGEFGIQ  
SGSQGCTLTSLPLDQOLENTADLPGLOGARLLVDDNETCKVLVQCCSGWGL  
NVSASVSGREALAQLRTKAHLREYFDVLLDQMPGTGQLAALKIKEPDLNHDILL  
IMLTGISNAPSRIARNAGIKRLAKPVGATYTLKATLADLAORGVSQVGTQFDVLV  
AQAPLSDFRILVAEDNSISTKVTGMLNKLINLOPDTASNGEALSAMKATQVLDVL  
MDCMEVLDGFSATQEAWEHQRPPTPVVALTATLHSEKREARLVGMDGHMAKP  
VELSQLRLIAYVWGERRRRQDGLPS"  
4760..5353  
/gene="PA4857"  
4760..5353  
/gene="PA4857"  
/codon\_start=1  
/transl\_table=11

```

/product="hypothetical protein"
/db_xref="GI:9951129"
/translation="MHLLSVYKLVLLVSPFFVLSVCFISLRSFGPGRDKRMARVVA
LAALIASVALYLFGRYIFILFEGITADFRIGAGSVLFIISALGMAQCAVQSDNVQOD
VTIVLPIDLPVGPCTIGALLVNGVCGWKAELISLVAIFLACLTLGLTLLSRIER
LLGQGLQIVSRMLGCLFVCAAAQIIVGLKSLFPG"
/length=7043
/genes="PA4858"
5778..7043
/genes="PA4858"
5778..7043
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08243.1"
/db_xref="GI:9951130"
/translation="MKRSLLKKAFTLSAFIASMGLSWSIOAETIKVIGLHLSGTMW
ISLTKDMALMTIDEINAKGVNGRKLKLEAVVDPASNPFLFAKARQLLTQDKVAV
FGWTSSRKSFLVPEELNGLLFYQVYEGEELS PNVTYGAAPNQQAIPAVIYLS
EDGGARRFFLLGTDYVYPTTKILRAF LSHKGVDDKDIQEVYTPFGHSIDYITLAD
IKKFAAGKTAIVSTVNGDSNVDPYKELANOCLEATEVPVVAFSVGEELRGIDTKPL
VGNLAAMNYSVSNPNYNEKFAVDKAYAKAKNLPNTGTATNDPMEATYVGIHWKQ
AVKAGSTEVDKVRKEMAGOSFKAPSGYTLVMDKSNHHLKHPVMIGEIQDDQEVVM
KTEPLRAQPSFPFGNEKKPEHALKSN"
/length=7826
/genes="PA4859"
7826..8779
/genes="PA4859"
7826..8779
/codon_start=1
/transl_table=11
/product="probable permease of ABC transporter"
/protein_id="AAG08244.1"
/db_xref="GI:9951131"
/translation="MRTAAATSLAQVRRLLVGDLLQAFSGLSGILLSLAAUGLAI
TYGLGVINMAHGEMLGAYSTYVQVLLQRLAPWELALYPLLALPVAFVCSAIGM
ALERTVIRHLYGRPLETLATGWSILVLIQVRLFGAONVEANPWLSSGIGIOLPN
LVLPWNRIVIGFALVLLTLTLLNRLGLNVRAYTQNRNMAACCGVPTGRVMDILA
FGSGSIGLGVALSQVNGVDPGLGGYIIDSFLVVLGVGLGAGSVLAAGLGLVA
NKILEPQIGAVLGKILILALVLFIQKRPQGLFALKGRVID"
/length=8776
/genes="PA4860"
8776..9855
/genes="PA4860"
8776..9855
/codon_start=1
/transl_table=11
/product="probable permease of ABC transporter"
/protein_id="AAG08245.1"
/db_xref="GI:9951132"
/translation="MNQPLVTLARAGRVTAALGAFAATLLVMPLLHLLPADSAL
HLSAYSLTUTGKLCYALVALDLYVNGAGLLSGHLFFLGGTAMGNTLMKQASG
DSLPAFTLSWNLPLPWAGTOHFLWALCLVLPVPGALAFAGFYAFKSRIRGVYFS
IMTOALTGAMLLFRNETGFGNGGFTDFRSILGFPITAPGTRAAFLATVALLAAS
LLGLWRLARSKEGRVITLALDAENRLMFCGYPGRYKLVYVLTSLVLCGLAGALFVPO
VGINSPSPTNSIEAAVAVVVLGGRGTLGLGAGVWNGAKSWFTVAFPEYWLFL
GALFIVVTLYLPRLGLLLRKEQ"
/length=9852
/genes="PA4861"
9852..10709
/genes="PA4861"
9852..10709
/codon_start=1
/transl_table=11
/product="probable ATP-binding component of ABC transporter"
/protein_id="AAG08246.1"
/db_xref="GI:9951133"
/translation="MRTTPHLMLEPAHPGLDLAGSRDAIGLGAATSGRLDARHGTI
LSLEDINVSDFGKALDRLTYIGVGLRCILGPNAGAKTLLMTITGTRPQSGTAY
FGDTLDRMSVQIAQIGRFRKPTVFLEALSVFENLELAQRADKSWASLARLD
GTQREIEVFLATIRLESQRPAAGLLSHGQKFLIEIGMLLVQEPOLLILLDEPVAGMT
DAETETAEFLSKARHSLMVMVHDGFGVGSIAHDHVTVLHQCHVLAEGSLAEVQADE
QVTEVYLR"
/length=1768
BASE COUNT 1768 a 3916 c 3416 g 1707 t
ORIGIN

```

```

Query Match          0.9%; Score 22; DB 1; Length 10807;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 gttcttcgacctgctgtagct 87
      |||||
Db 6704 gttcttcgacctgctgtagct 6683

RESULT 4
AC021564/c
LOCUS
DEFINITION
Homo sapiens clone RP11-772J6, WORKING DRAFT SEQUENCE, 17 unordered
pieces.
AC021564
VERSION
AC021564.3 GI:7417805
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 170970)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone Rpl1-772J6
Unpublished
REFERENCE
2 (bases 1 to 170970)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,I., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenesco,J.,
Ferrario,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severly,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6899714.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5687
Center clone name: 772-J_6
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160384 bases at least Q40
Consensus quality: 165238 bases at least Q30
Consensus quality: 167127 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 169370; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```





```

source
1. .30
/organism="Aspergillus oryzae"
/db_xref="taxon:5062" 9 t

BASE COUNT      12 a      3 c      6 g      9 t
ORIGIN

Query Match      0.9%; Score 21; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2030 ccataccttttcgataact 2050
|||||
Db 30 CCATACCTTTTCGTATAAGT 10

RESULT 7
AC011538
LOCUS      Homo sapiens 38627 bp DNA PRI 21-JUN-2001
DEFINITION Homo sapiens chromosome 19 clone LLNLF-252C1, complete sequence.
ACCESSION AC011538
VERSION AC011538.6 GI:14518401
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 38627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38627)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jun 21, 2001 this sequence version replaced gi:7711529.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .38627
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLF-252C1"

BASE COUNT      9713 a      7023 c      7436 g      14455 t
ORIGIN

Query Match      0.9%; Score 21; DB 9; Length 38627;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2196 tattgtatattcattaatgaa 2216
|||||
Db 9616 TATTGTATATTCATTAAATGAA 9636

RESULT 8
AC021090/c
LOCUS      Homo sapiens 40611 bp DNA PRI 27-MAY-2000
DEFINITION Homo sapiens chromosome 19 clone LLNLF-175G4, complete sequence.
ACCESSION AC021090
VERSION AC021090.5 GI:8099267

```

```

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 40611)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40611)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 27, 2000 this sequence version replaced gi:7711670.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .40611
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLF-175G4"

BASE COUNT      14376 a      7707 c      7409 g      11119 t
ORIGIN

Query Match      0.9%; Score 21; DB 9; Length 40611;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2196 tattgtatattcattaatgaa 2216
|||||
Db 17399 TATTGTATATTCATTAAATGAA 17379

RESULT 9
AC092279
LOCUS      Homo sapiens 105538 bp DNA HTG 03-JUL-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-2017D11, WORKING DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AC092279
VERSION AC092279.1 GI:14589468
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 105538)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 105538)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 634160, BC599184

```

Center clone name: CITB-HI\_2017D11  
-----  
Summary Statistics  
Consensus quality: 91971 bases at least Q40  
Consensus quality: 100707 bases at least Q30  
Consensus quality: 102669 bases at least Q20  
Estimated insert size: 104538; sum-of-contigs estimation  
Quality coverage: 5.05 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.76 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1048: contig of 1048 bp in length  
\* 1049  
\* 1148: gap of unknown length  
\* 1149  
\* 2182: contig of 1034 bp in length  
\* 2183  
\* 2282: gap of unknown length  
\* 2283  
\* 4691: contig of 2409 bp in length  
\* 4692  
\* 4791: gap of unknown length  
\* 4792  
\* 7286: contig of 2495 bp in length  
\* 7287  
\* 7386: gap of unknown length  
\* 7387  
\* 11789: contig of 4403 bp in length  
\* 11790  
\* 11899: gap of unknown length  
\* 11890  
\* 18634: contig of 6745 bp in length  
\* 18635  
\* 18734: gap of unknown length  
\* 18735  
\* 31948: contig of 13214 bp in length  
\* 31949  
\* 32048: gap of unknown length  
\* 32049  
\* 41163: contig of 9115 bp in length  
\* 41164  
\* 41263: gap of unknown length  
\* 41264  
\* 56571: contig of 15308 bp in length  
\* 56572  
\* 56671: gap of unknown length  
\* 56672  
\* 75326: contig of 18655 bp in length  
\* 75327  
\* 75426: gap of unknown length  
\* 75427  
\* 105538: contig of 30112 bp in length.

Location/Qualifiers  
1. 105538  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CTD-2017D11"  
/clone.lib="Caltech human BAC library D"  
BASE COUNT 28572 a 22456 c 22832 g 30654 t 1024 others  
ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 105538;  
Best local similarity 100.0%; Pred. No. 8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2196 tatgtatattcaataagaa 2216  
|||||  
Db 29427 TATGTATATTCATTAAAGAA 29447

RESULT 10  
AC091983/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-559F22, WORKING DRAFT  
SEQUENCE, 41 unordered pieces..  
ACCESSION AC091983  
VERSION AC091983.1 GI:14333919  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 153173)  
AUTHORS DOE Joint Genome Institute.

Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 153173)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 621163  
Center clone name: RPCI-11\_559F22  
-----  
Summary Statistics  
Consensus quality: 106792 bases at least Q40  
Consensus quality: 130951 bases at least Q30  
Consensus quality: 136968 bases at least Q20  
Estimated insert size: 196170; agarose-fp estimation  
Estimated insert size: 149173; sum-of-contigs estimation  
Quality coverage: 5.72 in Q20 bases; agarose-fp estimation  
Quality coverage: 7.52 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 41 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1113: contig of 1113 bp in length  
\* 1114  
\* 1213: gap of unknown length  
\* 1214  
\* 2311: contig of 1098 bp in length  
\* 2312  
\* 2411: gap of unknown length  
\* 2412  
\* 4000: contig of 1589 bp in length  
\* 4001  
\* 4100: gap of unknown length  
\* 4101  
\* 5397: contig of 1297 bp in length  
\* 5398  
\* 5497: gap of unknown length  
\* 5498  
\* 6591: contig of 1094 bp in length  
\* 6592  
\* 6692: gap of unknown length  
\* 6693  
\* 7877: contig of 1185 bp in length  
\* 7878  
\* 7976: gap of unknown length  
\* 7977  
\* 9394: contig of 1418 bp in length  
\* 9395  
\* 9494: gap of unknown length  
\* 9495  
\* 10965: contig of 1471 bp in length  
\* 10966  
\* 11065: gap of unknown length  
\* 11066  
\* 13041: contig of 1976 bp in length  
\* 13042  
\* 13141: gap of unknown length  
\* 13142  
\* 14376: contig of 1235 bp in length  
\* 14377  
\* 14476: gap of unknown length  
\* 14477  
\* 15688: contig of 1212 bp in length  
\* 15689  
\* 15788: gap of unknown length  
\* 15789  
\* 17784: contig of 1996 bp in length  
\* 17785  
\* 17884: gap of unknown length  
\* 17885  
\* 19277: contig of 1843 bp in length  
\* 19278  
\* 19827: gap of unknown length  
\* 19828  
\* 20855: contig of 1028 bp in length  
\* 20856  
\* 20955: gap of unknown length  
\* 20956  
\* 22120: contig of 1165 bp in length  
\* 22121  
\* 22220: gap of unknown length  
\* 22221  
\* 23448: contig of 1228 bp in length  
\* 23449  
\* 23548: gap of unknown length  
\* 23549  
\* 25675: contig of 2127 bp in length  
\* 25676  
\* 25775: gap of unknown length  
\* 25776  
\* 27096: contig of 1321 bp in length  
\* 27097  
\* 27196: gap of unknown length  
\* 27197  
\* 29338: contig of 2142 bp in length  
\* 29339  
\* 29438: gap of unknown length  
\* 30938: contig of 1500 bp in length  
\* 30939  
\* 31038: gap of unknown length  
\* 31039  
\* 32662: contig of 1624 bp in length

```

* 32663 32762: gap of unknown length
* 32763 34189: contig of 1427 bp in length
* 34190 34289: gap of unknown length
* 34290 35589: contig of 1300 bp in length
* 35590 35689: gap of unknown length
* 35690 37304: contig of 1615 bp in length
* 37305 37404: gap of unknown length
* 37405 39017: contig of 1613 bp in length
* 39018 39117: gap of unknown length
* 39118 40664: contig of 1547 bp in length
* 40665 40764: gap of unknown length
* 40765 42207: contig of 1443 bp in length
* 42208 43822: gap of unknown length
* 43823 43922: contig of 1515 bp in length
* 43923 46593: contig of 2671 bp in length
* 46594 46693: gap of unknown length
* 46694 49219: contig of 2526 bp in length
* 49220 49319: gap of unknown length
* 49320 51644: contig of 2325 bp in length
* 51645 51744: gap of unknown length
* 51745 54629: contig of 2885 bp in length
* 54630 54729: gap of unknown length
* 54730 56080: contig of 1351 bp in length
* 56081 56180: gap of unknown length
* 56181 59449: contig of 3269 bp in length
* 59450 59549: gap of unknown length
* 59550 64245: contig of 4696 bp in length
* 64246 64345: gap of unknown length
* 64346 69497: contig of 5152 bp in length
* 69498 69597: gap of unknown length
* 69598 72251: contig of 2654 bp in length
* 72252 72351: gap of unknown length
* 72352 77749: contig of 5398 bp in length
* 77750 77849: gap of unknown length
* 77850 87428: contig of 9479 bp in length
* 87429 87429: gap of unknown length
* 87429 103483: contig of 16055 bp in length
* 103484 103584: gap of unknown length
* 103584 153173: contig of 49590 bp in length.
FEATURES
    source
        location/Qualifiers
            1. 153173
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clone="RP11-559F22"
                /clone_lib="RPCI human BAC library 11"
BASE COUNT 45483 a 28901 c 29032 g 45744 t 4013 others
ORIGIN

Query Match      0.9%; Score 21; DB 2; Length 153173;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2196 tattgataatcattaatgaa 2216
Db 91857 TATTGATATTCAATTAATGAA 91837
|||||
RESULT 11
AC072046/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-450H5, WORKING DRAFT SEQUENCE, 11-NOV-2000
AC072046
VERSION AC072046.6 GI:11128191
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184606)

```

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Birmeah,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroli,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaik,F., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 184606)  
Worley,K.C.

Submitted (07-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 9, 2000 this sequence version replaced gi:9625287.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HBP4  
Center clone name: RP11-450H5  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 83% of reads  
Chemistry: Dye-terminator Big Dye; 83% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 163304 bases at least Q40  
Consensus quality: 174355 bases at least Q30  
Consensus quality: 178508 bases at least Q20  
Estimated insert size: 178126; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: this is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

Db 27702 GATATTCATATGCGAATAAT 27682			
RESULT 12		AC067865 200689 bp DNA HTG 25-JUN-2000	
LOCUS		Homo sapiens chromosome 2 clone RP11-314N9 map 2, WORKING DRAFT	
DEFINITION		SEQUENCE, 15 unordered pieces.	
AC067865		AC067865	
ACCESSION		AC067865.2 GI:87051161	
VERSION		HTG: HTGS_PHASE1; HTGS_DRAFT.	
KEYWORDS		human	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 200689)	
AUTHORS		Birren, B., Linton, L., Nusbaum, C. and Lander, E.	
TITLE		Homo sapiens chromosome 2, clone RP11-314N9	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 200689)	
AUTHORS		Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Bouknight, B., Brown, A., Burkett, G., Campolano, A., Castle, A., Choquet, F., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Hopland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meidram, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	
TITLE		Direct Submission	
JOURNAL		Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT		On Jun 25, 2000 this sequence version replaced gi:7651917. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html	
		----- Genome Center	
		Center: Whitehead Institute/ MIT Center for Genome Research	
		Center code: WIBR	
		Web site: http://www-seq.wi.mit.edu	
		Contact: sequence_submissions@genome.wi.mit.edu	
		----- Project Information	
		Center project name: L9065	
		Center clone name: 314_N_9	
		----- Summary Statistics	
		Sequencing vector: M13; M7815; 100% of reads	
		Chemistry: Dye-terminator Big Dye; 100% of reads	
		Assembly program: Phrap; version 0.960731	
		Consensus quality: 189802 bases at least Q40	
		Consensus quality: 195895 bases at least Q30	
		Consensus quality: 198051 bases at least Q20	
		Insert size: 194000; agarose-fp	
		Insert size: 199289; sum-of-contigs	
		Quality coverage: 4.5 in Q20 bases; agarose-fp	
		Quality coverage: 4.4 in Q20 bases; sum-of-contigs	
		-----	
		* NOTE: This is a 'working draft' sequence. It currently	
		* consists of 15 contigs. The true order of the pieces	
		* is not known and their order in this sequence record is	
		* arbitrary. Gaps between the contigs are represented as	

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* this record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1		21440:	contig of 21440 bp in length
21441		21540:	gap of unknown length
21541		40789:	contig of 19249 bp in length
40790		40889:	gap of unknown length
40890		57998:	contig of 17109 bp in length
57999		58098:	gap of unknown length
58099		72182:	contig of 14084 bp in length
72183		72282:	gap of unknown length
72283		81930:	contig of 9648 bp in length
81931		82030:	gap of unknown length
82031		90915:	contig of 8885 bp in length
90916		91015:	gap of unknown length
91016		95604:	contig of 4589 bp in length
95605		95704:	gap of unknown length
95705		104642:	contig of 8938 bp in length
104643		104742:	gap of unknown length
104743		113644:	contig of 8902 bp in length
113645		113744:	gap of unknown length
113745		120572:	contig of 6828 bp in length
120573		120672:	gap of unknown length
120673		127681:	contig of 7009 bp in length
127682		127781:	gap of unknown length
127782		13381:	contig of 6070 bp in length
13382		13391:	contig of unknown length
13392		139649:	contig of 3698 bp in length
139650		139749:	gap of unknown length
139750		145151:	contig of 5402 bp in length
145152		145251:	gap of unknown length
145252		150958:	contig of 5707 bp in length
150959		151058:	gap of unknown length
151059		155261:	contig of 4203 bp in length
155262		155361:	gap of unknown length
155362		159825:	contig of 4464 bp in length
159826		159945:	gap of unknown length
159926		164213:	contig of 4288 bp in length
164214		164313:	gap of unknown length
164314		167956:	contig of 3643 bp in length
167957		168056:	gap of unknown length
168057		171292:	contig of 3236 bp in length
171293		171392:	gap of unknown length
171393		174510:	contig of 3118 bp in length
174511		174610:	gap of unknown length
174611		176567:	contig of 1957 bp in length
176568		176667:	gap of unknown length
176669		178426:	contig of 1759 bp in length
178427		178526:	gap of unknown length
178527		180699:	contig of 2173 bp in length
180700		180799:	gap of unknown length
180800		182025:	contig of 1226 bp in length
182026		182125:	gap of unknown length
182126		183460:	contig of 1335 bp in length
183461		183560:	gap of unknown length
183561		184606:	contig of 1046 bp in length.

FEATURES  
Source  
1. 184606  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-450H5"  
BASE COUNT 60191 a 32101 c 30675 g 58997 t 2642 others  
ORIGIN  
Query Match 0.9% Score 21; DB 2; Length 184606;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 712 gatattcatatgtggaataat 732

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1975: contig of 1975 bp in length  
 1976 2075: gap of 100 bp  
 2076 4372: contig of 2297 bp in length  
 4373 4472: gap of 100 bp  
 4473 8606: contig of 4134 bp in length  
 8607 8706: gap of 100 bp  
 8707 16318: contig of 7612 bp in length  
 16319 16418: gap of 100 bp  
 16419 22960: contig of 6542 bp in length  
 22961 23060: gap of 100 bp  
 23061 32721: contig of 9661 bp in length  
 32722 32821: gap of 100 bp  
 32822 40636: contig of 7815 bp in length  
 40637 40736: gap of 100 bp  
 40737 54297: contig of 13561 bp in length  
 54298 54397: gap of 100 bp  
 54398 70328: contig of 15931 bp in length  
 70329 70428: gap of 100 bp  
 70429 87196: contig of 16768 bp in length  
 87197 87296: gap of 100 bp  
 87297 104624: contig of 17328 bp in length  
 104625 104724: gap of 100 bp  
 104725 125960: contig of 21236 bp in length  
 125961 126060: gap of 100 bp  
 126061 146883: contig of 20823 bp in length  
 146884 146983: gap of 100 bp  
 146984 174757: contig of 27774 bp in length  
 174758 174857: gap of 100 bp  
 174858 200689: contig of 25832 bp in length.

## FEATURES

Location/Qualifiers  
 1. .200689  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 1. .1975  
 /note="assembly\_fragment"  
 2076. .4372  
 /note="assembly\_fragment"  
 4473. .8606  
 /note="assembly\_fragment"  
 8707. .16318  
 /note="assembly\_fragment"  
 16419. .22960  
 /note="assembly\_fragment"  
 23061. .32721  
 /note="assembly\_fragment"  
 32822. .40636  
 /note="assembly\_fragment"  
 40737. .54297  
 /note="assembly\_fragment"  
 54398. .70328  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:right  
 70429. .87196  
 /note="assembly\_fragment"  
 87297. .104624  
 /note="assembly\_fragment"  
 104725. .125960  
 /note="assembly\_fragment"  
 126061. .146883  
 /note="assembly\_fragment"  
 146984. .174757  
 /note="assembly\_fragment"  
 174858. .200689  
 /note="assembly\_fragment"

BASE COUNT 68219 a 34336 c 34359 g 62374 t 1401 others  
 ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 200689;  
 Best local Similarity 100.0%; Pred. NO. 7.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 712 gatattcatgttggaataat 732  
 |||||

Db 199512 GATATTGATGCGAATAAT 199532

## RESULT 13

AC067900 201312 bp DNA HTG 09-MAY-2001  
 Homo sapiens chromosome 8 clone RP11-775M3 map 8, WORKING DRAFT  
 SEQUENCE, 5 unordered pieces.  
 AC067900  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 human.  
 AC067900.3 GI:12545341  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 201312)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## TITLE

Homo sapiens chromosome 8, clone RP11-775M3

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 201312)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campionlavik,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,J.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meidrim,J., Menues,L., Mihoval,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

## TITLE

Journal

## COMMENT

Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jan 26, 2001 this sequence version replaced gi:8138718.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project information  
 Center project name: L7524  
 Center clone name: 775\_M\_3  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 41% of reads  
 Sequencing vector: plasmid; n/a; 59% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 198330 bases at least Q40  
 Consensus quality: 199387 bases at least Q30

Consensus quality: 200109 bases at least Q20  
 Insert size: 210000; agarose-fp  
 Insert size: 200912; sum-of-contigs  
 Quality coverage: 9.1 in Q20 bases; agarose-fp  
 Quality coverage: 9.5 in Q20 ba.  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 77871: contig of 77871 bp in length  
 \* 77872 77971: gap of 100 bp  
 \* 77972 83159: contig of 5188 bp in length  
 \* 83160 83259: gap of 100 bp  
 \* 83260 115460: contig of 32201 bp in length  
 \* 115461 115560: gap of 100 bp  
 \* 115561 158995: contig of 43435 bp in length  
 \* 158996 159095: gap of 100 bp  
 \* 159096 201312: contig of 42217 bp in length.

## FEATURES

Location/Qualifiers  
 1..201312  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP11-775M3"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1..77871  
 /note="assembly\_fragment"  
 clone\_end:SF6  
 vector\_side:left  
 77972..83159  
 /note="assembly\_fragment"  
 83260..115460  
 /note="assembly\_fragment"  
 115561..158995  
 /note="assembly\_fragment"  
 159096..201312  
 /note="assembly\_fragment"  
 65566 a 34813 c 36561 g 63964 t 408 others

BASE COUNT 65566 a 34813 c 36561 g 63964 t 408 others

## ORIGIN

Query Match 0.9% Score 21; DB 2; Length 201312;  
 Best Local Similarity 100.0%; Pred. No. 7.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2222 ttgataacatgggtattaa 2242  
 |||||  
 Db 70363 TTGATAACATGGGATTAA 70383

## RESULT 14

AC073534  
 LOCUS AC073534 209910 bp DNA HTG 20-APR-2001  
 DEFINITION Homo sapiens chromosome 19 clone CTD-2027119, WORKING DRAFT  
 SEQUENCE, 74 unordered pieces.  
 AC073534  
 VERSION AC073534.3 GI:13699747  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 209910)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 19  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 209910)  
 AUTHORS DOE Joint Genome Institute.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
 Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Apr 20, 2001 this sequence version replaced gi:9554695.  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----

Project Information  
 Center Project Name: 638128, BC603152  
 Center clone name: CTFB-HI\_2027119  
 -----

## Summary Statistics

Consensus quality: 121533 bases at least Q40  
 Consensus quality: 142219 bases at least Q30  
 Consensus quality: 152387 bases at least Q20  
 Estimated insert size: 141340; agarose-fp estimation  
 Estimated insert size: 202610; sum-of-contigs estimation  
 Quality coverage: 9.39 in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.55 in Q20 bases; sum-of-contigs estimation.  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 74 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1303: contig of 1303 bp in length  
 \* 1304 1403: gap of unknown length  
 \* 1404 3247: contig of 1844 bp in length  
 \* 3248 3347: gap of unknown length  
 \* 3348 4368: contig of 1021 bp in length  
 \* 4369 4468: gap of unknown length  
 \* 4469 5629: contig of 1161 bp in length  
 \* 5630 5729: gap of unknown length  
 \* 5730 6752: contig of 1023 bp in length  
 \* 6753 6852: gap of unknown length  
 \* 6853 8296: contig of 1444 bp in length  
 \* 8297 8396: gap of unknown length  
 \* 8397 9437: contig of 1041 bp in length  
 \* 9438 9537: gap of unknown length  
 \* 9538 10687: contig of 1149 bp in length  
 \* 10687 10786: gap of unknown length  
 \* 10787 11792: contig of 1006 bp in length  
 \* 11793 11892: gap of unknown length  
 \* 11893 12911: contig of 1019 bp in length  
 \* 12912 13011: gap of unknown length  
 \* 13012 14060: contig of 1049 bp in length  
 \* 14061 14160: gap of unknown length  
 \* 14161 15213: contig of 1053 bp in length  
 \* 15214 15313: gap of unknown length  
 \* 15314 16503: contig of 1190 bp in length  
 \* 16504 16603: gap of unknown length  
 \* 16604 17729: contig of 1126 bp in length  
 \* 17730 17829: gap of unknown length  
 \* 17830 18997: contig of 1168 bp in length  
 \* 18998 20118: contig of 1021 bp in length  
 \* 20119 20218: gap of unknown length  
 \* 20219 21253: contig of 1035 bp in length  
 \* 21254 21353: gap of unknown length  
 \* 21354 22435: contig of 1082 bp in length  
 \* 22436 22535: gap of unknown length  
 \* 22536 23539: contig of 1004 bp in length  
 \* 23540 23639: gap of unknown length  
 \* 23640 24791: contig of 1152 bp in length  
 \* 24792 24891: gap of unknown length  
 \* 24892 26023: contig of 1132 bp in length  
 \* 26024 26123: gap of unknown length  
 \* 26124 27261: contig of 1138 bp in length  
 \* 27262 27361: gap of unknown length

under conditions of nitrogen deprivation, which is the normal situation for *B. cinerea* during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

## FEATURES

source Location/Qualifiers  
1..240  
/organism="Botryotinia fuckeliana"  
/strain="T4"  
/db\_xref="taxon:40559"  
/note="Genoscope sequence ID : W13D061"

BASE COUNT  
74 a 50 c 65 g 51 t

ORIGIN

Query Match 0.9%; Score 20; DB 8; Length 240;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 648 ctgaaggagacaaggtcgaga 667  
|||||  
Db 102 CTTGAAGGACAAAGGTCGAGA 121

Search completed: December 4, 2001, 18:48:10  
Job time: 11394 sec

